# ABSTRACT

| | |
|---|---|
| Title of dissertation: | ESSAYS ON THE ECONOMICS OF SKILLS |
| | Fernando Andrés Saltiel<br>Doctor of Philosophy, 2019 |
| Dissertation directed by: | Professor Sergio Urzúa<br>Department of Economics |

In this dissertation, I examine the importance of specific components of the skill vector in affecting outcomes across various settings. In particular, I first consider the importance of non-cognitive skills in higher education in the United States, both in explaining academic undermatch, but also showing their importance towards successful degree completion. In the Chilean context, I consider how early-life math skills affect the likelihood of reaching the top of the income distribution, partly through leading to employment in higher-quality firms. The last chapter of my dissertation presents a discrete choice model of college majors, in which I consider how non-cognitive skills contribute to the gender gap in STEM majors in the United States. In particular, I document the importance of mathematical self-efficacy as an important driver of the gender gap in STEM.

In Chapter 2, I analyze the importance of non-cognitive skills in the context of higher education. Using longitudinal data for the United States, I first find that students with higher non-cognitive skills are more likely to enroll in higher-quality four-year colleges. Furthermore, students who have been previously characterized

as "under-matched" in higher education have significantly lower non-cognitive skills than students with equivalent test scores. While enrollment is the first step towards higher education completion, a burgeoning literature has documented falling completion rates among enrollees. In this context, I find that for both two-year enrollees as well as those in four-year colleges of varying qualities, non-cognitive skills are strong predictors of subsequent college completion.

Chapter 3, written in collaboration with Sergio Urzua, estimates the returns to skills in the labor market by taking advantage of three administrative data sources. We first test for non-linearities in these returns and find that the returns to mathematical skills are highly non-linear, with math skill 'superstars' far outearning other high math scorers. High math-skilled workers not only complete more years of education, but graduate from higher quality universities and earn higher-paying degrees. We further examine the role of firms as a mediator of the returns to skills, a dimension not previously explored in the literature. We find that high-skilled workers match to high-paying firms immediately upon labor market entry. We conduct a decomposition to examine the separate contribution of education and firms in mediating the returns to skills, and find that worker-firm matching explains almost half of the estimated returns.

Chapter 4 studies the relationship between pre-college skills and the gender gap in STEM majors. I expand upon the analysis in the first two chapters, by introducing structure to students' human capital investment decisions using a discrete choice model of college major choices. I implement the model using longitudinal data for the United States and consider students' initial and final major choices in a context where college students sort into majors based on observed characteristics and unobserved ability. More specifically, I distinguish observed test scores from latent ability. I find that math test scores significantly overstate gender gaps in math problem solving ability. Math problem solving ability strongly predicts STEM enrollment and completion for men and women. I further explore the importance of math self-efficacy, which captures students' beliefs about their ability to perform math-related tasks. Math self-efficacy raises both men's and women's probability of enrolling in a STEM major. Math self-efficacy also plays a critical role in explaining decisions to drop out of STEM majors for women, but not for men. The correlation between the two math ability components is higher for men than for women, indicating a relative shortfall of high-achieving women who are confident in their math ability. Lastly, I estimate the returns to STEM enrollment and completion and find large returns for high math ability women. These findings suggest that well-focused math self-efficacy interventions could boost women's STEM participation and graduation rates. Further, given the high returns to a STEM major for high math ability women, such interventions also could improve women's labor market outcomes.

# ESSAYS ON THE ECONOMICS OF SKILLS

by

Fernando Andrés Saltiel

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Sergio Urzúa Chair/Advisor
Professor Judith Hellerstein
Professor Melissa Kearney
Professor Susan W. Parker
Professor Nolan Pope

# Dedication

To my parents, Olga and Gustavo, who always supported me.

A Luciana, por estar siempre.

# Acknowledgments

I am greatly indebted to Sergio Urzúa for the time he devoted to advising my research and for his continued support of my research ideas. His work and dedication showed me the path to exploring questions I have always been interested in, while following a rigorous empirical approach. His support in pushing me to apply to grants, conferences and academic seminars early on in my career has been invaluable. Melissa Kearney's belief in me as an undergraduate student sparked my interest in research. Her commitment and passion to rigorous research which could shape the policy discussion showed me how academic research could impact the world in a variety of ways. She has been a terrific advisor and mentor during these years.

I am extremely grateful to other faculty members, as well. Judy Hellerstein instilled in me a passion for rigorous research in labor economics. Lesley Turner's support during my first year in the program was invaluable. Jessica Goldberg showed me the value of mentoring and advising undergraduate students. I also thank Nolan Pope and Susan Parker for agreeing to participate in the dissertation committee.

I am grateful to my classmates and friends, including Alejandro Graziano, Mateo Uribe-Castro, Santiago Velez-Ferro, Juan Agustín Echenique, Matthew Staiger, Cody Tuttle, Prateik Dalmia, Sai Luo, Camila Galindo-Pardo, Rodrigo Heresi, Cristian Sanchez, and Riley Wilson. Their company and extended research discussions helped me enormously throughout my graduate studies and made significant contributions to my dissertation.

My parents and my sister supported me unconditionally throughout this long and arduous process. They always believed in me and taught me the value of an education from early on. Luciana has been there for me every day during my graduate studies. I am forever grateful for her patience and support through the difficult times in these years. I am greatly looking forward to starting the next chapter of our lives together.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

An extensive literature has shown cognitive and non-cognitive skills to be strong predictors of various outcomes in adulthood, including employment, occupational choices, wages, as well as non-market outcomes (Heckman et al., 2006; Lochner, 2011). Nonetheless, the recent growth in college enrollment rates has brought increased prominence to specific human capital investment decisions. For instance, Altonji et al. (2012) have shown that the difference in wage returns across some majors exceeds the college wage premium. In this context, understanding the skill components which drive specific human capital investment decisions is of critical importance to shape future skill development policies.

In this dissertation, I examine the importance of specific components of the skill vector in affecting outcomes across different contexts. In particular, I focus my attention on understanding how early-life skills affect human capital investment decisions, and consider how such decisions subsequently affect labor market outcomes. The thesis is comprised of three chapters. The first chapter presents reduced form evidence on how a lack of non-cognitive skills both leads to academic undermatch in higher education in the United States, but also to a lower likelihood of successfully completing a higher education degree. In the second chapter, I consider how early-life math skills affect the likelihood of reaching the top of the income distribution, partly through leading to employment in higher-quality firms, presenting a decomposition to understand the various channels through which early life skills affect labor market outcomes. In the last chapter, I introduce a discrete choice model to carefully model how students' select their college majors. I examine how non-cognitive skills contribute to the gender gap in STEM majors in the United States, focusing on the importance of mathematical self-efficacy.

This dissertation contributes to the existing literature on the economics of education in several ways. First, I present extensive evidence of how individuals sort into college majors based on their pre-college skills. By showing that mathematical self-efficacy plays a critical role in the decision to pursue a STEM degree, I argue that boosting women's self-efficacy in math could lead to narrower gender gaps in this field. Using data from Chile, I further show how the returns to early-life math skills follow a non-linear pattern, leading to the presence of math superstars. By showing that these superstars partly emerge through matching to high-quality firms, I contribute to the small, but growing literature on firm-worker matching. Lastly, I show that students who are traditionally considered undermatched in higher education have lower non-cognitive skills vis-a-vis their well-matched counterparts. While the existing literature has argued that undermatch is largely driven by information frictions, these results indicate that non-cognitive skill interventions may lead to improved matches in higher education, as well.

The second chapter examines how non-cognitive skills affect college enrollment decisions and degree completion across colleges of varying qualities in the United

States. In particular, I note that non-cognitive skills may play an important role in enrollment decisions trough the application process, which requires students to identify the colleges they plan to apply to, take the ACT or the SAT, request high school transcripts and recommendation letters from teachers, and complete college essays, among other tasks (Avery and Kane, 2004; Carrell and Sacerdote, 2017). Moreover, as previous work has found that an important driver of mismatch is the application process, students who are lacking in grit and perseverance may fail to complete applications or miss critical deadlines and thus enroll in a lower-quality college than one they could have been admitted to (Hoxby and Avery, 2013; Dillon and Smith, 2017). Furthermore, non-cognitive skills may also help students after enrolling in college, where they often face various difficulties, including living far from family, adapting to a new environment and facing challenging courses, among others.

I take advantage of data from the Educational Longitudinal Study of 2002 (ELS), which includes detailed information on multiple measures of pre-college preparation, non-cognitive skill measures and detailed information on students' enrollment and subsequent completion across higher education institutions of varying qualities. I estimate various reduced form regressions and first find that non-cognitive ability positively predicts enrollment in higher-quality institutions. I also show that this skill dimension is an important predictor of "academic undermatch," where highly-qualified students enroll in lower-quality colleges. In the fact of declining completion rates among college enrollees, I also show that a one standard deviation increase in non-cognitive skills among four-year college enrollees is associated with an increased likelihood of completing a Bachelor's degree by age 26 by 4.5 percentage points. These results suggest that non-cognitive skill development policies may be a plausible alternative for improving outcomes in higher education.

In the third chapter, written in collaboration with Sergio Urzua, we estimate the labor market returns to mathematical and language skills in the labor market, examine potential non-linearities in these returns and explore how educational attainment and firm-quality matching mediates the returns to ability. This question has gained prominence in the context of skill-biased technological change, which delivers increased wages for higher-skilled workers, relative to their less skilled counterparts (Katz et al., 1998; Acemoglu, 2002; Card and DiNardo, 2002).

Taking advantage of three administrative data sources from Chile, we examine the returns to a nationally-administered standardized math and reading exam given to $10^{th}$ graders, who are matched to their subsequent educational attainment and labor market outcomes in their early 30s. We find large returns to the math exam in the labor market, which follows a non-linear pattern, such that moving from the $50^{th}$ percentile of the math test score distribution to the $85^{th}$ percentile results in a 30 percent wage increase, the same returns are observed by moving from the $95^{th}$ to the $99^{th}$ percentile. In examining the mechanisms behind this result, we find that math superstars attain more years of education relative to their less-skilled peers, but also attend the highest-quality colleges and graduate from the highest-paying majors. We extend this analysis to consider how firm-worker matching mediates the returns to skills. In this context, we find that the math superstars match with high-quality firms immediately upon labor market entry. To discern the importance of education and firm-worker matching in explaining the returns to skills, we estimate an augmented human capital equation and find that the estimated returns to math ability fall from 22% to 7-9% once detailed measures of education *and* firm quality are included. We lastly estimate a Gelbach (2016) decomposition and find that firms explain half of the aggregate returns to math test scores, highlighting an important mechanism through which skills deliver improved labor market outcomes.

The last chapter examines the factors driving the gender gap in STEM majors by estimating a discrete choice model of students' college major choices. The gender gap in STEM majors carries significant policy relevance, as women make-up just one fourth of recent graduates in math-intensive STEM majors in the United States (Kahn and Ginther, 2017), and these majors are among the highest-paying college degrees. Moreover, various colleges across the countries have begun implementing policies aimed at increasing STEM participation, yet the factors driving students enrollment/completion decisions across college majors are not well understood. Following from the literature which as documented sizable gender gaps in math test scores can explain the difference in STEM participation (Turner and Bowen, 1999; Xie and Shauman, 2003; Dickson, 2010; Riegle-Crumb et al., 2012; Justman and Méndez, 2018), I analyze whether pre-college skills can account for part of the gender gap in STEM participation. In this context, I extend the existing literature by distinguishing math test scores from underlying math problem solving ability and considering the importance of relevant non-cognitive skills, focusing on the on the role of mathematical self-efficacy, which measures an individual's perceived ability to perform math-related tasks.

To understand students' enrollment and completion patterns given their pre-college ability, I estimate a sequential model of college major choices. The model builds on Heckman et al. (2016), Heckman et al. (2018) and Humphries et al. (2017) and assumes that at each stage, individual decisions and labor market outcomes are a function of observed characteristics and latent math and reading ability, which differ from observed test scores. I take advantage of Educational Longitudinal Study of 2002 (ELS) data, which follows a nationally-representative cohort of $10^{th}$ graders through age 26. ELS data includes detailed information on multiple measures of math test scores, math class GPA, math self-efficacy measures, detailed information college major choices and early-career labor market outcomes.

I first find that the gender gap in unobserved math problem solving ability is 40 percent smaller than the 0.30 standard deviation gap in math test scores. This result is explained by the fact that women outperform men in math high school courses, and GPA partly reflects underlying math ability. I show that women have lower math self-efficacy than men, but also document a lower correlation between math ability and math self-efficacy for women than for men. Since both math problem solving ability and self-efficacy are strong predictors of STEM enrollment for both men and women, the relative lack of women at the top of the joint skill distribution reduces their initial participation in math-intensive STEM majors. Given that 60% of men initially enrolled in STEM majors end up graduating relative to just 45% of women, I examine whether math self-efficacy can partly explain differential completion rates. In this context, I show that women's lower math self-efficacy accounts for 20% of the gender gap in STEM completion rates among STEM enrollees.

All in all, this dissertation contributes to an extensive literature examining the factors driving human capital investment decisions as well as the consequences associated with those decisions. By highlighting how different components of the skills vector affect these decisions, this thesis sheds light on how early-life skill development can play a critical role in changing individuals' educational attainment as well as their labor market outcomes.

# Chapter 2: Gritting it Out: The Importance of Non-Cognitive Skills in Higher Education

## 2.1 Introduction

The sizable increase in college participation in recent decades has been largely concentrated in two-year and less selective four-year institutions (Bound et al., 2010). At the same time, an extensive literature has shown that students enrolled in higher-quality colleges are more likely to graduate Brewer et al. (1999); Long (2008); Goodman et al. (2017); Dillon and Smith (2018), and to earn higher wages (Black and Smith, 2006; Hoekstra, 2009; Zimmerman, 2014). As a result, understanding the factors driving students' enrollment decisions across colleges of varying qualities is of critical importance, especially in light of falling college completion rates. In this context, while previous work has considered the importance of family background Hoxby and Avery (2013); Smith et al. (2013), information frictions Hoxby and Turner (2015); Pallais (2015); Lincove and Cortes (2016) and cognitive skills Light and Strayer (2000); Kinsler and Pavan (2011), this literature has not considered the importance of non-cognitive skills in determining enrollment decisions.

Why might non-cognitive skills play an important role in higher education?

An important factor in the determining enrollment decisions in the United States is the application process, which requires students to identify the colleges they plan to apply to, take the ACT or the SAT, request high school transcripts and recommendation letters from teachers, and complete college essays, among other tasks (Avery and Kane, 2004; Carrell and Sacerdote, 2017). In fact, as Hoxby and Avery (2013) and Dillon and Smith (2017) have found that an important driver of mismatch is the application process, students who are lacking in grit and perseverance may fail to complete applications or miss critical deadlines and thus enroll in a lower-quality college than one they could have been admitted to. Furthermore, non-cognitive skills may also help students after enrolling in college, where they often face various difficulties, including living far from family, adapting to a new environment and facing challenging courses, among others. In this context, Oreopoulos and Ford (2016) document that less conscientious students are are more likely to under-perform in college, indicating that non-cognitive skills may affect students' likelihood of successfully completing a college degree.

In this chapter, I analyze how non-cognitive skills affect students' enrollment decisions across colleges of varying qualities and examine how this skill dimension helps them in successfully completing a college degree. While previous work has shown the positive impact of "soft" skills on individuals' final educational attainment Heckman et al. (2006, 2018), the literature has not yet analyzed the importance of this skill dimension across the college quality distribution nor has it explored its differential impacts at enrollment and graduation. I examine this question using data from the Educational Longitudinal Study of 2002 (ELS), which follows a nationally-representative cohort of $10^{th}$ graders through age 26. ELS data includes detailed information on multiple measures of pre-college preparation, various measures of non-cognitive skills and detailed information on students' enrollment and subsequent completion across higher education institutions of varying qualities.

I first estimate a multinomial logit model of college enrollment and find that non-cognitive ability positively predicts enrollment in higher-quality institutions. For instance, a one standard deviation increase in non-cognitive skills raises the likelihood of enrollment in a highly-selective college by 3.5 percentage points. This effect is 40% as large as an equivalent increase in a math test score and larger than the corresponding impact for the reading exam. I find larger impacts for men than for women, indicating that this skill dimension may play a role in differential gender enrollment rates (Goldin et al., 2006). Since non-cognitive skills are important for the enrollment decision, I also examine whether this component of skills plays a role in explaining "academic undermatch," where highly-qualified students enroll in lower-quality colleges. I find that 42% of students in the math test score decile enroll in institutions below the highly-selective group and this group trails their "well-matched" counterparts by 0.28 standard deviations in a composite index of non-cognitive skills. Furthermore, among students in the top two math test score quintiles, those enrolled in selective or highly-selective colleges outscore their undermatched counterparts by 0.31 SDs in the non-cognitive skill index. While previous work had suggested that undermatch could be partially remedied through information-based interventions Hoxby and Avery (2013), these results indicate that a relative lack of non-cognitive skills may be leading highly-qualified students to enroll in lower-quality colleges.

Given the recent decrease in completion rates among college enrollees, I also analyze the importance of non-cognitive skills for subsequent degree completion. Fitting in with findings by Beattie et al. (2017), a one standard deviation increase in the non-cognitive skill index among four-year college enrollees increases the likelihood of completing a Bachelor's degree by age 26 by 4.5 percentage points, equaling half of the effect of the math test score. The magnitude of the effect is similar across men and women. Furthermore, this result holds across the college quality distribution. Moreover, among two-year college enrollees, non-cognitive skills have a similar-sized effect on the likelihood of completing either a two- or a four-year degree by age 26, and the magnitude of this effect is two-thirds as large as that of the math test score. All in all, given the importance of non-cognitive skills in driving enrollment choices and college graduation, the malleability of this skill component through late adolescence indicates that non-cognitive skill development policies may be a plausible alternative for improving outcomes in higher education (Kautz et al., 2014).

The rest of the chapter is organized as follows. Section 2 describes the data sources and presents summary statistics. In Section 3, I present evidence on how pre-college test scores and non-cognitive skills shape initial enrollment choices. I also present evidence on the importance of non-cognitive skills in academic mismatch. In Section 4, I explore how these dimensions affect students' final educational attainment by age 26. Finally, in Section 5, I present conclusions and final remarks.

## 2.2 Data Sources and Summary Statistics

### 2.2.1 Data Sources

This chapter uses longitudinal data from the Educational Longitudinal Survey (ELS) of 2002.[1] The ELS is a nationally-representative survey of 16,200 $10^{th}$ grade students in 2002 who were interviewed, along with their parents and teachers, in the initial year, and in 2004, 2006, and 2012, when respondents had turned 26/27 years old. The first two surveys include detailed information on students' individual characteristics, including their race and gender, and family characteristics, including family composition, parents' educational attainment, labor market outcomes, total family income and region of residence. Furthermore, ELS data includes students' performance on a mathematics and reading test developed by the Department of Education in $10^{th}$ grade, and a follow-up math exam in $12^{th}$ grade.

ELS data includes various questions capturing respondents' non-cognitive skills, as well, which were measured in the baseline survey. I focus on three measures, which were directly constructed by ELS staff using exploratory factor analysis on the responses to these questions. The first variable captures a student's expectations of success in academic learning (control expectation scale), which follows from their responses to four different statements.[2] The second measure captures students' per-

[1]The ELS is a part of the National Center for Education Statistics' program which includes three earlier longitudinal studies of high school students in the United States: the National Longitudinal Study of the High School Class of 1972, the High School and Beyond Longitudinal study of 1980, and the National Education Longitudinal Study of 1988.

[2]Students are prompted with the following statements: "I can learn something really hard," "I can get no bad grades if I decide to," "I can get no problems wrong if I decide to," "I can learn something well if I want to," and their responses are graded on a four-point Likert scale.

ceived effort and persistence when facing difficulties, constructed from the responses to five statements (action control), as well.[3] This question is most closely related to grit, a non-cognitive measure capturing persistence on tasks (Duckworth, Peterson, Matthews, and Kelly, Duckworth et al.)). The last variable captures students' motivation to perform well academically in order to reach external goals like future job opportunities or financial security (instrumental motivation). I examine students' performance in each of these variables, but I also construct an index of non-cognitive skills, which averages out their responses across the three measures.[4]

Critical to the analysis of sorting into colleges, ELS data includes detailed information on respondents' progression through higher education, including students' enrollment status in 2006 and final educational attainment in 2012 (age 26). I first classify students by the level of the their academic institution, which can either be a two- or four-year college. Four-year college enrollees are further classified by their college's selectivity, following an ELS-provided classification in which colleges pertain to one of three mutually exclusive categories: inclusive, moderately-selective and highly-selective colleges. The classification follows from the 2005 Carnegie Classification System, which is based on the distribution on the $25^{th}$ percentile of admitted students' ACT/SAT test scores. Highly selective colleges are those in the

[3]The statements are as follows: "I remember most important things when I study," "I work as hard as possible when I study," "I keep studying even if the material is difficult," "I do my best to learn," "I put forth my best effort when studying."

[4]Since Humphries and Kosse (2017) argue that papers in this literature may reach different conclusions depending on the measures underlying the construction of non-cognitive skill indices, I construct a non-cognitive index which follows from principal component analysis of the questions/statements answered by students. The results are unchanged when using this index and they are available upon request. These concerns are common to other commonly-used longitudinal data sets, such as the NLS72, NLSY79, NELS88, since there are no comparable non-cognitive skill measures across these surveys.

top quintile of this measure, moderately-selective institutions are in the fourth and third quintile, and inclusive four-year colleges are those in the bottom two quintiles.[5] Other papers in this literature use different definitions of selectivity, yet these measures depend directly on the data source used for empirical analysis. I note that the measure used in this chapter matches exactly the one used by two other papers using ELS data (Reardon, 2011).[6]

### 2.2.2 Summary Statistics

In the first panel of Table 4.1, I present descriptive statistics on the sample used in this chapter (Column 1). Women comprise 53% of the sample, 63% of students are white and the share of black and Asian students reaches almost 10%.[7] The majority of students in the sample come from two parent families, the average surveyed parent has almost completed 15 years of schooling, and average family income for students in the sample exceeds $55,000.[8] In terms of students' educational attainment by age 26, just 45% of the sample has completed at least a two-year degree, with 38% of students having completed a four-year degree and the remaining students finishing an Associate's degree.

[5] I include open admissions institutions in the same category as inclusive colleges. The results are robust to analyzing these two sets of institutions separately.

[6] Bound et al. (2010) use NLS72 and NELS:88 data. They classify college selectivity following *US News and World Report* rankings. Using NLSY data, Light and Strayer (2000) construct an index of college quality from Barron's Profiles of American Colleges. Meanwhile, Black and Smith (2004) and Dillon and Smith (2017) construct an index of latent college quality using different measures of institutional quality available in the IPEDS data.

[7] The over-representation of women is explained by the fact that women are more likely to have completed the baseline and follow-up surveys. In the empirical results, I use sample weights to account for survey non-response.

[8] Family income is reported to pertain to one of 13 mutually exclusive bins and I assign family income to equal the mid-point of the reported bin. I note that this is a noisy measure of family resources, though similar issues have been found in NLSY97 data (Dillon and Smith, 2017).

I further examine students' baseline characteristics and in final outcomes by their initial enrollment decision. The characteristics of students not enrolled in college by age 20/21 are shown in Column 2.[9] This group is largely comprised of males, fitting in with Goldin et al. (2006)'s finding of higher enrollment rates for women. Moreover, these students are less likely to come from two-parent families compared to the rest of the sample and their families earn lower annual incomes and have less educated parents, on average. In terms of their academic achievement, this group lags far behind the rest of the sample, with average math and reading scores 0.5 standard deviations below their enrolled counterparts.[10] This difference is also present in the three measures of non-cognitive skills, in the range of 0.3 standard deviations, as well as in the composite index of non-cognitive skills. Unsurprisingly, this group is the least likely to have completed a higher education degree by age 26, with just 0.9% of students graduating with a four-year college degree. The two-year college enrollee sample accounts for 23% of the full sample and these students come from higher-educated and higher-income families relative to the non-enrolled group. The two-year sample outpaces non-enrollees in academic preparation and in non-cognitive skills, with an average 0.16 SD difference in average math and reading scores and a 0.13 standard deviation difference in the composite non-cognitive skill index. Two-year enrollees have a higher likelihood of completing a higher-education degree by age 26, although only 40% of students in the sample do so.

Four-year enrollees come from higher-income and higher-educated families rel-

[9]This category includes students enrolled in $<$ 2-year institutions as well as non-degree seekers.

[10]Test score and non-cognitive skill measures are standardized to be mean zero and standard deviation one.

ative to students in the other two groups. Furthermore, they outperform the rest of the sample in terms of academic preparation, outscoring two-year enrollees by 0.75 standard deviations in the math test and by 0.72 SDs in the English test. These differences also appear in the non-cognitive dimension, with four-year enrollees outscoring those in the other groups across the three observed measures as well as in the composite index. As 68% of students complete a four-year degree by age 26, this group has the highest educational attainment in the final follow-up survey. Nonetheless, there are sizable differences within this group, depending on the quality of the college attended. I explore these differences in Panel B. First, about 40% of college enrollees attend a highly-selective college, with an equal share attending moderately selective colleges. There are small differences across the three groups in reported family income, though students in highly-selective colleges are more likely to come from a two-parent family and have more educated parents. There are larger differences in academic preparation, though, as the highly-selective group outscores those in inclusive institutions by 0.9 standard deviations in the math test and the selective group by 0.5 SDs. These differences are larger than those between the open-enrollment group and non-enrollees, highlighting the importance of considering four-year college quality when examining enrollment decisions. These differences are also present across the skill distribution, as the first panel in Figure 2.1 shows that math test score distribution of students in highly-selective colleges dominates that of students in the other categories.[11] Those in highly-selective col-

[11]While students enrolled in highly-selective colleges make-up 18 percent of the sample, they account for 54% of students in the top math test score decile.

leges outpace the rest of the sample in the non-cognitive skill dimension as well, as shown in the bottom rows of Table 4.1 and in the second panel of Figure 2.1. Finally, confirming previous findings by Bound et al. (2010) and by Dillon and Smith (2018), students in higher-quality institutions are far more likely to have completed a four-year degree by age 26. While I have so far documented sizable differences in academic preparation and in non-cognitive skills across students' enrollment decisions, the descriptive analysis does not control for the importance of other variables in affecting enrollment choices. I next present reduced-form evidence on the importance of test scores and non-cognitive skills at enrollment.

## 2.3 Enrollment Decision

To explore how test scores, non-cognitive skills and background characteristics affect students' enrollment decisions by age 20, I first estimate a multinomial logit as follows:

$$Y_i^E = \beta_0 + \beta_1 X_i + \beta_2 \theta_i^C + \beta_3 \theta_i^{NC} + \varepsilon_i \tag{2.1}$$

The outcome variable $Y_i^E$ denotes the student $i$'s initial enrollment decision by age 20. This variable takes on five values, including non-enrollment, two-year enrollment, four-year non-selective, selective and highly-selective enrollment. $X_i$ includes race/ethnicity, gender, family composition and income, along with parents' education. $\theta_i^C$ includes the baseline math and reading test scores and $\theta_i^{NC}$ captures the non-cognitive skill index. While equation (2.1) provides descriptive, rather than causal, evidence of the determinants of enrollment choices, I note that this analysis is a first approximation of the importance of non-cognitive skills in students initial higher education choices.

Table 2.2 presents the marginal effects of each variable on the enrollment decision. Two-year college enrollment is the excluded category. Men are less likely to enroll in moderately- and highly-selective colleges and more likely to not have enrolled in college by age 20, confirming patterns found in the descriptive analysis. Students coming from two-parent households and those with more educated parents are more likely to have enrolled in a highly-selective four-year institution, yet reported family income does not a significant impact at enrollment. Unsurprisingly, conditional on background characteristics, academic preparation still has a sizable impact on students' college quality at enrollment. For instance, a one standard deviation increase in the English test score increases the probability of enrollment in a highly selective college, relative to two-year college enrollment, by five percentage points. Meanwhile, an equivalent increase in the math test score raises the relative likelihood of highly-selective enrollment by 8.5 percentage points, while reducing the probability of not enrolling in college by a similar magnitude. These results fit in with previous work by Light and Strayer (2000), Kinsler and Pavan (2011) and Lovenheim and Reynolds (2013), who use NLSY79 and NLSY97 data and find that high-AFQT students are more likely to enroll in high-quality four-year colleges.

However, previous research on this topic has not considered the importance of non-cognitive skills at enrollment. In this context, I find that conditional on academic preparation and background characteristics, non-cognitive skills have an important effect on students' college enrollment decision by age 20: a one standard deviation increase in the non-cognitive skill index raises the relative probability of enrollment in a moderately-selective college by 1.5 percentage points and in a highly-selective institution by 3.3 percentage points, respectively, while significantly reducing the likelihood of not enrolling in college by age 20. The magnitude of the marginal effects of non-cognitive skills on enrollment in a high-quality institution is similar to that of the English test score, and reaches almost 40% of the estimated effect of the math test score.

In Appendix Table A.1.1, I examine the contribution of each separate component of the non-cognitive skill index. I find that a one standard deviation increase in the action control component, which is the most closely related to "grit", increases the probability of enrollment in a highly-selective college by 2.3 percentage points, relative to starting in a two-year college. The instrumental motivation measure, which captures a student's motivation to achieve future goals, also has a positive impact on starting in a high-quality college, yet the magnitude is significantly smaller. An equivalent increase in a student's expectations of academic success has no discernible effect on enrollment choices. I complement this analysis by exploring whether sorting patterns differ by gender in Appendix Table A.1.2, as male college enrollment rates are far behind those of women (Goldin et al., 2006). For both men and women, students with a higher non-cognitive skill endowment are more likely to enroll in a highly-selective institution relative to a two-year college. However, the estimated impact of soft skills is half as large as that of the math test score for men, whereas the ratio of these coefficients is below 30% for women.

As a result, non-cognitive skill development strategies may help in increasing male college enrollment rates, but also in driving them to higher-quality institutions. I note that while the literature examining the determinants of enrollment decisions has often equated cognitive test scores, like the AFQT, with students' pre-college ability, the results shown so far indicate that non-cognitive skill measures should be taken into account when analyzing sorting patterns by pre-college ability. I next examine whether a lack of non-cognitive skills leads high-achieving students to be more likely to "undermatch" at enrollment.

### 2.3.1 Academic Undermatch

An extensive literature has analyzed the drivers of 'academic undermatch,' which occurs when a student's academic qualifications would allow her to enroll in a higher-quality college compared to their current institution. This phenomenon may negatively impact undermatched students, as previous research has found that higher-quality colleges increase the likelihood of degree completion Brewer et al. (1999); Light and Strayer (2000); Dillon and Smith (2018) and improve labor market outcomes (Black and Smith, 2006; Bowen et al., 2009; Hoekstra, 2009). In the United States, Smith et al. (2013) found that 40 percent of high school students undermatch in their postsecondary choice, and Bowen et al. (2009) reach similar conclusions using data from North Carolina. Various papers in this literature have shown that undermatch is more prevalent among students from low-income families, students residing in non-urban areas as well as those not close to a well-matched public university (Smith et al., 2013; Hoxby and Avery, 2013; Dillon and Smith, 2017).

On the other hand, this literature has not considered the importance of non-cognitive skills in determining academic mismatch. This skill dimension may play a role in enrollment decisions, especially in light of the complexity of the college application process. As noted above, the structure of the application process implies that students who are less likely to persist in the face of difficulty (low action control), may fail to execute each step of this process and thus end up enrolling in colleges with lower admission requirements. Similarly, students with lower instrumental motivation may be *ceteris paribus* less motivated to pursue a college degree, and thus choose to apply to a less-selective college with a simpler application process, or enroll in a non-selective two- or four-year institution.

An important consideration in this literature is the exact definition of under-match, which varies across papers depending on data availability and on the research question. For instance, Dillon and Smith (2017) classify students as undermatched if the difference between ASVAB percentile and their college quality percentile is below a certain threshold. Meanwhile, Roderick et al. (2008) and Smith et al. (2013) consider students to be undermatched if they enroll in a lower-quality institution than one in which they are likely to be admitted. Since the main goal of this chapter is to understand the role of non-cognitive skills in higher education, rather than the nature or prevalence of undermatch, I define undermatch using three different proximate measures. The first measure focuses on students in the top math test score decile, who given their academic performance should be able to enroll in a highly-selective institution. These students are defined to be "under-matched," if they have not enrolled in one of these colleges by age 20. Among those in the top decile, 42% of students are classified as undermatched. The second measure also focus on students not enrolled in highly-selective colleges, but extends the sample to include students in the top test score *quintile*: 52% of these students are "undermatched," as they have not enrolled in a highly-selective institution. Finally, I examine the prevalence of undermatch among students in the top two test score deciles, considering them to be undermatched if they are either enrolled in an inclusive four-year institution or enrolled in two-year college. In this case, 37% of students are undermatched.

In Figure 2.2, I first present graphical evidence on the importance of non-cognitive skills in academic undermatch. The first panel shows that among students in the top test score decile, the distribution of the non-cognitive skill index for undermatched students is first-order stochastically dominated by that of top achievers enrolled in highly-selective institutions. In fact, these students outpace their 'undermatched' counterparts by 0.31 standard deviations in the index, by 0.32 SDs in the action control measure and by 0.31 standard deviations in instrumental motivation dimension. In the second panel, I find similar patterns, as undermatched students in the top two test score deciles trail their well-matched counterparts in the non-cognitive skill index. While the results in Figure 2.2 offer preliminary evidence of the importance of soft skills in determining undermatch, it does not control for family and individual characteristics. To examine the importance of non-cognitive skills in academic undermatch, I thus estimate the following linear probability model:

$$Undermatch_{i,k} = \alpha_0 + \alpha_1 X_i + \alpha_2 \theta_i^C + \alpha_3 \theta_i^{NC} + \epsilon_i \tag{2.2}$$

where $Undermatch_{i,k}$ represents one of the three measures of undermatch defined above.[12] As in equation (2.1), I include individual and family characteristics as well as measures of test score performance and non-cognitive skills.

I present the results in Table 2.3.[13] The first column examines the drivers of undermatch among students in the top math test score decile. As in Hoxby and Avery (2013) and Dillon and Smith (2017), I find that high-achieving students with less educated parents and those living in non-urban areas are less likely to have enrolled in a highly-selective college by age 20. Furthermore, conditional on reaching the top test score decile, an additional increase in both the math and the English test score decreases the likelihood of academic undermatch. Non-cognitive skills also play an important role, as a one standard deviation increase in the non-cognitive index decreases the probability of undermatch by 6.5 percentage points, or 15 percent of the baseline undermatch rate in this group. In Appendix Table A.1.3, I examine the separate contribution of each component of the non-cognitive skill, and find that the estimated effect is largely explained by the action control component, thus indicating that lower-"grit" high-achievers are more likely to undermatch in higher education. In the second column, I examine the prevalence of undermatch among students in the top math quintile. I find similar determinants effects as in the first column, yet the estimated impact of the math and English test scores is larger than for students in the top decile. Similarly, a one SD increase in the

[12]Since each definition of undermatch applies for a sub-sample of students, given their math test score performance, I estimate equation (2.2) only including students who are potentially undermatched under each definition.

[13]The results presented in Table 2.3 include students who had not enrolled in college by age 20 as 'undermatched.' In Appendix Table A.1.4, I exclude non-enrollees and re-estimate equation (2.2) across the three definitions of undermatch. The results are unchanged.

soft-skill index reduces the prevalence of undermatch by 7.5 percentage points, or 14 percent of the baseline rate. I lastly examine the determinants of undermatch among students in the top two test score deciles. Once again, I find that higher test score performers and those with a higher stock of non-cognitive skills are less likely to have undermatched in higher education.

Previous work had found that information-based interventions could help in reducing undermatch for high-achieving students (Hoxby and Turner, 2015). On the other hand, the results presented in this section suggest that a subset of high-achieving students may fail to enroll in high-quality colleges in part due to a shortfall in the non-cognitive dimension. As a result, students may fail to "grit out" the complex application process required for acceptance at a highly-selective college, and choose to enroll at open-enrollment or two-year colleges, instead. These results suggest that policies aimed at fostering non-cognitive skills may help in reducing undermatch. Nonetheless, while improved enrollment choices for students is an important first step towards increasing educational attainment, a sizable share of college enrollees fail to complete a degree (Bound et al., 2010). Therefore, I next explore how pre-college skills affect degree completion rates.

## 2.4 College Completion

As noted above, college enrollment does not necessarily lead to degree completion. In fact, NCES (2018) note that among four-year college enrollees in the 2009 starting cohort, only 59 percent had completed a degree within four years of enrollment.

In this context, previous work has found that enrolled students from lower-income families are less likely to complete a degree (Bailey and Dynarski, 2011; Stinebrickner and Stinebrickner, 2003), as are under-represented minority students across all enrollment levels (NCES, 2018). The importance of academic preparation in persistence has also been previously documented, with Bound et al. (2010) finding higher completion rates among high math-achievers and Agan (2013) finding that college dropouts have the lower AFQT scores than both Associate- and Bachelor-degree recipients. At the same time, as students face repeated challenges in higher education, including the transitioning from high school to college, making friends in a new context and persisting through difficult courses, they may need to rely on their non-cognitive skills in order to complete a degree. To explore the importance of these components in degree attainment, I estimate the following linear probability model:

$$G_{i,k} = \gamma_0 + \gamma_1 X_i + \gamma_2 \theta_i^C + \gamma_3 \theta_i^{NC} + e_i \tag{2.3}$$

where $G_{i,k}$ indicates whether student $i$ has completed degree $k$ by age 26, which includes two- and four-year degree completion, as well as graduating with a degree across each level of institutional selectivity.[14] As in equations (2.1) and (2.2), I include individual and family characteristics as well as measures of test score performance and non-cognitive skills.

[14]Given the prevalence of college transfers among four-year enrollees (Andrews et al. 2014), I examine whether students complete a degree in their initial enrollment level or complete a four-year degree at all.

I present the results of the determinants of Bachelor's degree attainment for four-year enrollees in Table 2.4. In the first column I include both males and females and find, as in Bowen et al. (2009), that students with more educated parents and those in two-parent households are more likely to successfully complete a four-year degree. Fitting in with citetbound2010college, I further find that better academically-prepared students are more likely to graduate, as a one standard deviation increase in the math test score is associated with a 9.3 percentage point increase in predicted completion rates, with a smaller predicted effect of the reading test score. At the same time, the non-cognitive index indicates that soft skills play an important role in determining degree completion, as a one SD increase in this index raises the likelihood of completion by 4.6 percentage points — equaling half of the magnitude of the math test score and exceeding the relative importance of reading skills. These results indicate students who are lacking in non-cognitive skills are paying a "double penalty" in higher education, first at college enrollment and subsequently at completion. The last two columns examine the drivers of completion separately for men and women, respectively. As in Appendix Table A.1.2, I find that non-cognitive skills play a larger role in determining college completion for men than for women, which suggests that this skill component could play a role in explaining lower college persistence among males (Conger and Long, 2010).

In Table 2.5, I expand upon this analysis by exploring whether the determinants of college completion differ across levels of college quality. The first column analyzes the factors driving college completion for students in open enrollment or inclusive institution — only 46% of students in these colleges complete a degree by age 26. While academic preparation, as measured by the math and reading test scores, plays an important role in leading to degree receipt, I also find that a one standard deviation increases in the non-cognitive index raises the likelihood of completion by 5.1 percentage points. I find similar effects for students enrolled in selective colleges, though for these students the relative importance of non-cognitive skills is larger, reaching 70% of the estimated effect of the math test score. On the other hand, for students enrolled in highly selective colleges, non-cognitive skills do not affect the likelihood of subsequent degree completion, unlike for their counterparts at less selective institutions. These results, combined with those presented in Section 3, suggest that part of the gap in graduation rates across college selectivity levels may be explained by a lack of non-cognitive skills among students enrolled in less selective institutions.

Lastly, as discussed above, degree completion rates are lowest among students enrolled in two-year institutions. In Table 2.6, I examine the factors affecting two-year and four-year degree receipt by age 26 among these students. I first note that women are more likely to complete any higher education degree, as are Asian students and those coming from families with more educated parents. As with four-year college enrollees, I find that both academic preparation and non-cognitive skills affect the likelihood of degree receipt. Nonetheless, there are significant differences on the relative importance of these skill components by the type of degree attained. While for two-year degree completion, the estimated effect of a one SD increase in the math test score is twice as large as one in the non-cognitive index, the relative effect of these components on four-year completion is not statistically different. Recall that since just 22% of two-year enrollees end up completing a bachelor's degree by age 26, the combination of the results presented in Tables 4 through 6 indicate that a boost in non-cognitive skills could significantly increase degree completion rates both for two-year enrollees but also for those who start in less selective four-year institutions.

## 2.5 Conclusion

While students enrolled in higher-quality colleges are more likely to complete a four-year degree and earn higher wages, the recent increase in college participation has been concentrated in lower-quality institutions. In this context, an extensive literature has examined the factors driving academic mismatch, where highly qualified students enroll in lower-quality institutions than they otherwise could, finding higher mismatch rates among lower-income students and those living in non-urban areas. In this chapter, I have considered the role of non-cognitive skills in higher education, aiming to understand their importance in driving enrollment decisions and academic mismatch. Across these two dimensions, I have found that students with a higher stock of non-cognitive skills are more likely to enroll in higher-quality colleges. Moreover, lower non-cognitive-skilled students are more likely to have undermatched in college, despite being high academic achievers. Since college enrollment requires students to sort through the various steps involved in the college application process, it is possible that less gritty students may instead choose to enroll in lower-quality colleges with simpler application procedures.

Recent work has also shown a significant decline in completion rates among college enrollees. As a result, I have also examined whether non-cognitive skills can predict subsequent completion, and found that for students in two-year institutions as well as in less-selective four-year colleges, this skill dimension is a strong predictor of degree attainment. As Kautz et al. (2014), among others, have shown that non-cognitive skills are malleable through adolescence, these results suggest that interventions aimed at fostering non-cognitive could lead to improved outcomes in higher education, both by reducing mismatch and by increasing college completion rates. While this chapter has shown descriptive evidence on the importance of non-cognitive skills in higher education, I have not yet considered how these skills affect college major choices as well as labor market outcomes. I address this issue in Chapter 4, where I estimate a discrete choice model of college major choices in the U.S., using the same dataset as in this chapter.

## 2.6 Tables and Figures

**Table 2.1:** Summary Statistics by Initial Enrollment Decision

**Panel A.** Summary Statistics by Enrollment Level

| | Full Sample | Not Enrolled | 2-Year | 4-Year |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Male | 0.467 | 0.550*** | 0.434 | 0.431 |
| White | 0.634 | 0.581 | 0.609 | 0.679*** |
| Black | 0.098 | 0.122*** | 0.092 | 0.085 |
| Asian | 0.097 | 0.072*** | 0.096 | 0.112** |
| Both Parents | 0.778 | 0.723*** | 0.760 | 0.819*** |
| Family Income | 10.916 | 10.771*** | 10.987 | 10.973 |
| Parents' Education | 14.833 | 13.917*** | 14.324 | 15.633*** |
| Math Test Score | 0.000 | -0.485*** | -0.318 | 0.446*** |
| English Test Score | 0.000 | -0.467*** | -0.300 | 0.427*** |
| Control Expectation | 0.000 | -0.286*** | -0.157 | 0.249*** |
| Instrumental Motivation | 0.000 | -0.221*** | -0.107 | 0.186*** |
| Action Control | 0.000 | -0.235*** | -0.131 | 0.206*** |
| Non-Cognitive Index | 0.000 | -0.279*** | -0.148 | 0.241*** |
| $\geq$Two-Year Graduate | 0.449 | 0.031*** | 0.397 | 0.731*** |
| Four-Year Graduate | 0.378 | 0.009*** | 0.224 | 0.677*** |
| Observations | 9,180 | 2,720 | 2,050 | 4,420 |

**Panel B.** Summary Statistics by Four-Year College Quality

| | Inclusive | Selective | Highly Selective |
|---|---|---|---|
| | (1) | (2) | (3) |
| Male | 0.427 | 0.417 | 0.450** |
| White | 0.550*** | 0.733 | 0.690*** |
| Black | 0.171*** | 0.077 | 0.046*** |
| Asian | 0.078 | 0.084 | 0.163*** |
| Both Parents | 0.764*** | 0.814 | 0.856*** |
| Family Income | 10.954** | 11.104 | 10.838 |
| Parents' Education | 14.830*** | 15.402 | 16.340*** |
| Math Test Score | -0.065*** | 0.347 | 0.842*** |
| English Test Score | -0.029*** | 0.356 | 0.762*** |
| Control Expectation | 0.048*** | 0.184 | 0.433*** |
| Instrumental Motivation | 0.043*** | 0.147 | 0.310*** |
| Action Control | 0.027*** | 0.140 | 0.379*** |
| Non-Cognitive Index | 0.045*** | 0.177 | 0.422*** |
| $\geq$Two-Year Graduate | 0.584*** | 0.715 | 0.832*** |
| Four-Year Graduate | 0.458*** | 0.661 | 0.815*** |
| Observations | 930 | 1,840 | 1,650 |

Source: Educational Longitudinal Study of 2002. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. In the first panel, the stars follow from a t-test of not enrolled and 4-year enrollees against those in two-year colleges, respectively. In the second panel, the stars follow from a t-test of students in inclusive colleges and highly-selective ones to those in selective universities, respectively.

**Table 2.2:** Enrollment Decision

| | Not Enrolled | 4-Yr Inclusive | 4-Yr Selective | 4-Yr Highly Selective |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Math Test Score | -0.0375*** | -0.00299 | 0.0376*** | 0.0852*** |
| | (0.00646) | (0.00468) | (0.00611) | (0.00551) |
| English Test Score | -0.0308*** | 0.00406 | 0.0204*** | 0.0501*** |
| | (0.00635) | (0.00463) | (0.00597) | (0.00528) |
| Non-Cognitive Index | -0.00932* | 0.00197 | 0.0151*** | 0.0324*** |
| | (0.00441) | (0.00321) | (0.00412) | (0.00357) |
| Male | -0.0320*** | -0.000962 | -0.0565*** | -0.0289*** |
| | (0.00878) | (0.00634) | (0.00799) | (0.00663) |
| White | -0.0151 | -0.0343*** | 0.0567*** | 0.0243* |
| | (0.0118) | (0.00861) | (0.0130) | (0.0113) |
| Black | -0.0655*** | 0.0377*** | 0.0570** | 0.0232 |
| | (0.0171) | (0.0107) | (0.0184) | (0.0174) |
| Asian | -0.0125 | -0.0298 | 0.0437* | 0.0711*** |
| | (0.0250) | (0.0181) | (0.0222) | (0.0169) |
| Both Parents | -0.0123 | -0.00429 | 0.00958 | 0.0232** |
| | (0.0105) | (0.00758) | (0.0101) | (0.00886) |
| Family Income | 0.0140*** | 0.00467 | 0.00518 | -0.00519** |
| | (0.00385) | (0.00260) | (0.00270) | (0.00180) |
| Parents' Education | -0.00823*** | 0.000960 | 0.00993*** | 0.0212*** |
| | (0.00187) | (0.00134) | (0.00172) | (0.00147) |
| Urban | -0.0252* | 0.0197** | 0.00430 | 0.0368*** |
| | (0.0103) | (0.00698) | (0.00905) | (0.00724) |
| Observations | | | 9,180 | |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table 2.2 presents the estimated marginal effects from a multinomial logit regression, as in equation (2.1), examining the determinants of initial enrollment decisions. The omitted category is two-year college enrollment.

**Table 2.3:** Determinants of Academic Undermatch

| | Top Test Score Decile | Top Test Score Quintile | Top Two Quintiles |
|---|---|---|---|
| | (1) | (2) | (3) |
| Math Test Score | -0.128* | -0.190*** | -0.160*** |
| | (0.0529) | (0.0304) | (0.0170) |
| English Test Score | -0.132** | -0.139*** | -0.0999*** |
| | (0.0452) | (0.0255) | (0.0145) |
| Non-Cognitive Index | -0.0664*** | -0.0753*** | -0.0556*** |
| | (0.0181) | (0.0121) | (0.00827) |
| Male | 0.0650* | 0.0632** | 0.146*** |
| | (0.0321) | (0.0221) | (0.0154) |
| White | 0.00255 | -0.0142 | -0.0929*** |
| | (0.0604) | (0.0416) | (0.0253) |
| Black | 0.206 | 0.160 | -0.0533 |
| | (0.167) | (0.0879) | (0.0483) |
| Asian | -0.141 | -0.105 | -0.102* |
| | (0.0817) | (0.0592) | (0.0410) |
| Both Parents | -0.129** | -0.0501 | -0.0264 |
| | (0.0432) | (0.0302) | (0.0203) |
| Family Income | 0.0137 | 0.00933 | 0.000135 |
| | (0.00747) | (0.00575) | (0.00440) |
| Parents' Education | -0.0532*** | -0.0453*** | -0.0395*** |
| | (0.00752) | (0.00504) | (0.00341) |
| Urban | -0.0921** | -0.0879*** | -0.0542** |
| | (0.0346) | (0.0250) | (0.0176) |
| Constant | 1.682*** | 1.660*** | 1.314*** |
| | (0.180) | (0.115) | (0.0748) |
| Observations | 920 | 1840 | 3670 |
| $R^2$ | 0.142 | 0.146 | 0.143 |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table 2.3 presents the estimated results from equation (2.2). Equation (2.2) estimates the determinants of academic undermatch for students at different points of the math test score distribution. In the first two columns, students are classified to undermatch if they are not enrolled in a highly-selective college. In the third column, undermatched students are those who have not enrolled in a selective or highly-selective institution by age 20.

**Table 2.4:** Determinants of Bachelor's Degree Completion Among Four-Year Enrollees

| | Four-Year Enrollees (1) | Males (2) | Females (3) |
|---|---|---|---|
| Math Test Score | 0.0925*** | 0.0811*** | 0.0984*** |
| | (0.0112) | (0.0172) | (0.0147) |
| English Test Score | 0.0297** | 0.0170 | 0.0452** |
| | (0.0108) | (0.0162) | (0.0147) |
| Non-Cognitive Index | 0.0459*** | 0.0514*** | 0.0414*** |
| | (0.00738) | (0.0114) | (0.00974) |
| Male | -0.0523*** | | |
| | (0.0140) | | |
| White | 0.0379 | 0.0698* | 0.0184 |
| | (0.0221) | (0.0353) | (0.0285) |
| Black | -0.0257 | 0.0352 | -0.0652 |
| | (0.0303) | (0.0484) | (0.0389) |
| Asian | 0.0681 | 0.115* | 0.0370 |
| | (0.0357) | (0.0545) | (0.0476) |
| Both Parents | 0.0437* | 0.0355 | 0.0487* |
| | (0.0177) | (0.0279) | (0.0229) |
| Family Income | -0.00211 | -0.000391 | -0.00320 |
| | (0.00394) | (0.00619) | (0.00510) |
| Parents' Education | 0.0276*** | 0.0310*** | 0.0250*** |
| | (0.00306) | (0.00485) | (0.00394) |
| Urban | 0.0137 | -0.0326 | 0.0486* |
| | (0.0151) | (0.0232) | (0.0200) |
| Constant | 0.140* | 0.0148 | 0.191* |
| | (0.0661) | (0.106) | (0.0847) |
| Observations | 4420 | 1910 | 2510 |
| $R^2$ | 0.109 | 0.096 | 0.123 |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table 2.4 presents the estimated results from equation (2.3). Equation (2.3) estimates the determinants of four-year degree completion by age 26 for students enrolled in four-year college by age 20. The first column includes all four-year enrollees, the whereas the next two columns divide the sample by gender.

**Table 2.5:** Determinants of Bachelor's Degree Completion Among Four-Year Enrollees

| | Inclusive/Open Enrollees | Selective College Enrolees | | Highly-Selective College Enrolees | |
|---|---|---|---|---|---|
| | Four-Year Grad. | Four-Year Grad. | >=Selective Grad. | Four-Year Graduate | Highly Selective Grad. |
| | (1) | (2) | (3) | (4) | (5) |
| Male | -0.0350 | -0.0577* | -0.0280 | -0.0431* | -0.0159 |
| | (0.0318) | (0.0229) | (0.0237) | (0.0197) | (0.0232) |
| White | 0.0419 | -0.00335 | -0.00798 | -0.0107 | -0.0208 |
| | (0.0416) | (0.0381) | (0.0395) | (0.0354) | (0.0416) |
| Black | 0.0398 | -0.0861 | -0.0699 | -0.0452 | -0.0624 |
| | (0.0506) | (0.0529) | (0.0548) | (0.0567) | (0.0666) |
| Asian | 0.222* | 0.0491 | -0.00255 | -0.0752 | -0.0506 |
| | (0.0885) | (0.0632) | (0.0655) | (0.0482) | (0.0567) |
| Both Parents | 0.0991** | 0.0152 | 0.00566 | 0.00965 | -0.00638 |
| | (0.0365) | (0.0286) | (0.0296) | (0.0273) | (0.0320) |
| Family Income | -0.0231 | 0.00225 | 0.0115 | -0.000292 | -0.00363 |
| | (0.0133) | (0.00724) | (0.00750) | (0.00440) | (0.00517) |
| Parents' Education | 0.0334*** | 0.0217*** | 0.0178*** | 0.0152*** | 0.0215*** |
| | (0.00672) | (0.00499) | (0.00517) | (0.00461) | (0.00541) |
| Urban | 0.0917** | 0.00129 | 0.00890 | -0.0407 | -0.0138 |
| | (0.0335) | (0.0257) | (0.0266) | (0.0208) | (0.0245) |
| Math Test Score | 0.0968*** | 0.0661*** | 0.0599** | 0.0476** | 0.0433* |
| | (0.0240) | (0.0182) | (0.0188) | (0.0176) | (0.0207) |
| English Test Score | 0.0460* | -0.00631 | 0.00328 | 0.0168 | 0.0498** |
| | (0.0231) | (0.0175) | (0.0181) | (0.0164) | (0.0193) |
| Non-Cognitive Index | 0.0513** | 0.0458*** | 0.0458*** | 0.0151 | 0.00272 |
| | (0.0162) | (0.0120) | (0.0124) | (0.0109) | (0.0128) |
| Constant | 0.0794 | 0.282* | 0.178 | 0.541*** | 0.336** |
| | (0.167) | (0.113) | (0.117) | (0.0968) | (0.114) |
| Observations | 930 | 1,840 | 1,840 | 1,650 | 1,650 |
| $R^2$ | 0.139 | 0.039 | 0.031 | 0.032 | 0.035 |

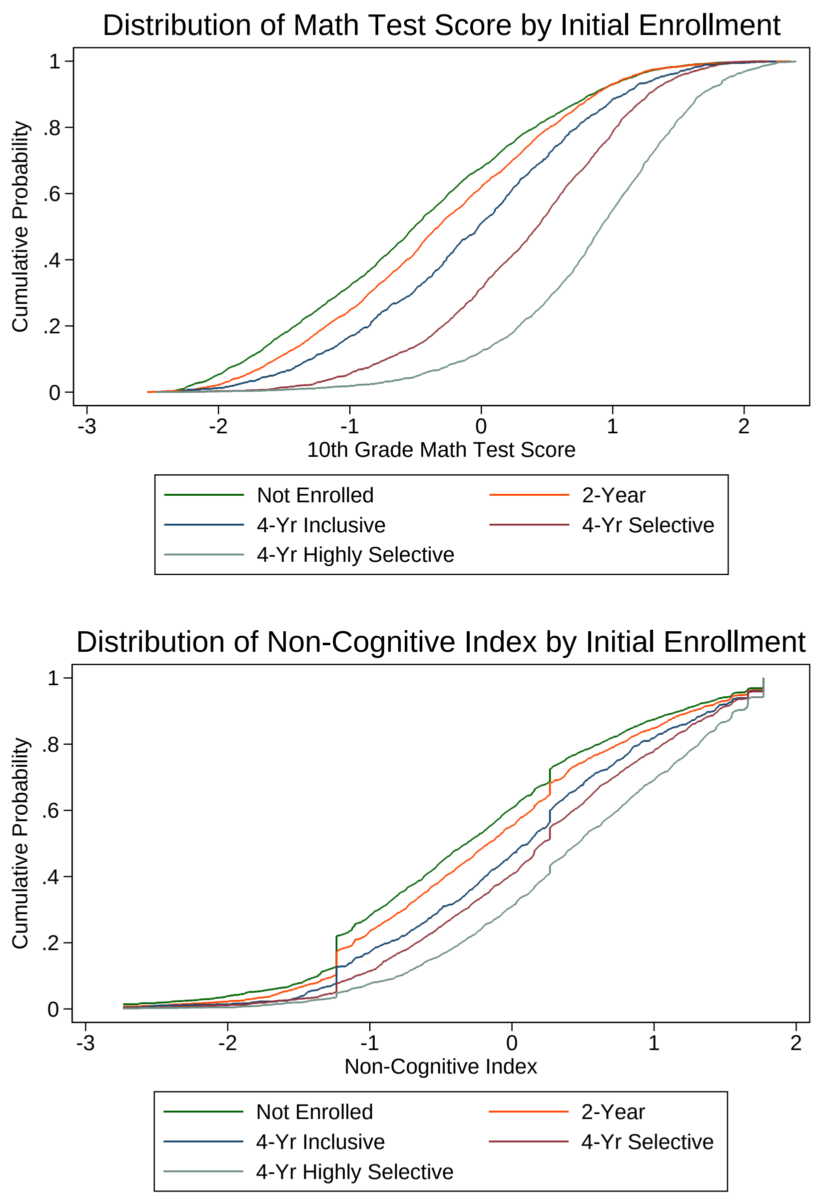
Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table 2.5 presents the estimated results from equation (2.3). Equation (2.3) estimates the determinants of four-year degree completion by age 26 for students enrolled in four-year college by age 20, separated by initial four-year college quality. The first column includes four-year enrollees in inclusive/open enrollment colleges. The second and third columns focus on those in selective colleges. The last two columns only include students in highly-selective institutions.

**Table 2.6:** Determinants of Bachelor's Degree Completion Among Two-Year Enrollees

| | (1)<br>Higher Education Graduate | (2)<br>Four-Year Graduate |
|---|---|---|
| Male | -0.0671** | -0.0802*** |
| | (0.0220) | (0.0185) |
| White | 0.0567* | 0.0294 |
| | (0.0285) | (0.0240) |
| Black | 0.00868 | -0.0196 |
| | (0.0414) | (0.0348) |
| Asian | 0.128* | 0.105* |
| | (0.0604) | (0.0507) |
| Both Parents | 0.0341 | 0.0138 |
| | (0.0257) | (0.0216) |
| Family Income | 0.0139 | 0.00500 |
| | (0.0110) | (0.00923) |
| Parents' Education | 0.0138** | 0.0168*** |
| | (0.00470) | (0.00395) |
| Urban | -0.0133 | 0.0317 |
| | (0.0261) | (0.0219) |
| Math Test Score | 0.0661*** | 0.0478*** |
| | (0.0163) | (0.0137) |
| English Test Score | 0.0158 | 0.0223 |
| | (0.0154) | (0.0129) |
| Non-Cognitive Index | 0.0341** | 0.0499*** |
| | (0.0107) | (0.00898) |
| Constant | 0.0412 | -0.0498 |
| | (0.126) | (0.106) |
| Observations | 2050 | 2050 |
| $R^2$ | 0.052 | 0.066 |

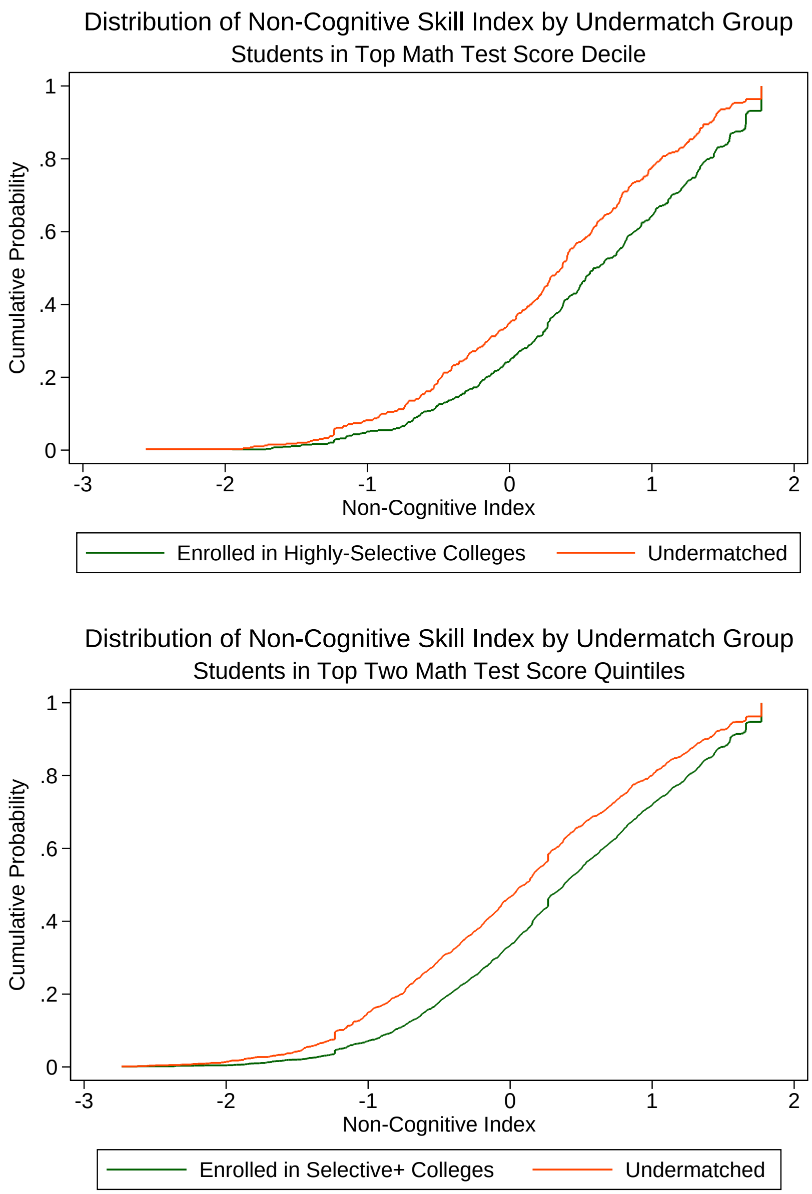
Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table 2.6 presents the estimated results from equation (2.3) for two-year college enrollees. The first column presents the probability of completing either a two- or a four-year degree by age 26.

**Figure 2.1:** Math Test Score and Non-Cognitive Skills by Initial Choice




Source: Educational Longitudinal Study of 2002. The first panel examines the distribution of the $10^{th}$ grade math test score by initial enrollment decisions. The second panel presents the distribution of the non-cognitive skill index by students' initial enrollment choice.

**Figure 2.2:** Non-Cognitive Skills and Undermatch in Higher Education




Source: Educational Longitudinal Study of 2002. The first panel examines the distribution of the non-cognitive index among top test score decile performers by their 'Undermatch' status. The second panel examines the distribution of the non-cognitive index among students in the top two math test score deciles, by their 'Undermatch' status. The definition of undermatch is described in Section 3.

# Chapter 3: Firms as Mediators of the Returns to Skills

*Note: This chapter of the dissertation is coauthored with Sergio Urzua.*

## 3.1 Introduction

In recent decades, the advent of skill-biased technological change has brought increased attention to the importance of workers' skills, with an early literature defining workers' skill levels by their educational attainment (Katz et al., 1998; Acemoglu, 2002; Card and DiNardo, 2002). However, recent work has highlighted the distinction between education and pre-labor market skills, where high-skilled individuals attain more years of education, resulting in human capital accumulation mediating the returns to skills (Heckman et al., 2006, 2018; Herrnstein and Murray, 1994; Blau and Kahn, 2005; Lindqvist and Vestman, 2011; Deming, 2017). Nevertheless, the bulk of this literature has estimated the returns to skill in regressions where ability enters as a linear term. As a result, these papers have so far overlooked potential non-linearities in the returns to skills, thus missing the existence and the returns to being a skill "superstar" (Rosen, 1981).

While education has been considered as a mediator of the returns to skills, the importance of firms as a conduit for these returns has not received extensive attention, despite their importance for affecting labor market outcomes (Abowd et al., 1999; Card et al., 2013, 2018). While previous research has found positive assortative matching between high-quality workers and high-quality firms Card et al. (2013, 2018), the worker "quality" measured in these papers is identified from observed labor market outcomes, and cannot be directly tied to measures of pre-labor market skill measures. Nonetheless, these findings suggest that an additional channel through which skills may result in higher wages may be through high-skilled workers matching to higher-quality firms, either immediately upon labor market entry or incrementally as workers acquire labor market experience.

In this chapter, we estimate the labor market returns to mathematical and language skills in the labor market, examine potential non-linearities in these returns and explore how educational attainment and firm-quality matching mediates the returns to ability.[1] To answer this question, we take advantage of three administrative data sources from Chile. First, we use test score data from a nationally-administered standardized math and reading exam given to $10^{th}$ graders in 2001 and in 2003, comprising a nationally-representative sample of the 1985 and 1987 birth cohorts. We match these students to administrative data on their high school records and their higher-education enrollment and graduation records in 2005-2017. These data sources include detailed information on degrees attained, the fields of these degrees and measures of university quality. Lastly, we match these individuals to administrative matched employee-employer data over the 2002-2016 period. This

[1] Throughout this chapter, we use the terms *skills* and *ability* interchangeably. Unlike Chapter 2, as we do not have access to information on students' non-cognitive skills, we abstract away from considering this skill dimension.

data includes monthly frequency of employment spells for workers in the formal sector, and we are able match over 240,000 workers from the test-taking sample. As a result, we are able to follow workers from labor market entry through ages 29-31.

We find large returns to mathematical ability in the labor market: a one standard deviation increase in $10^{th}$ grade math test scores increases monthly earnings by 20%, whereas the equivalent returns to language ability are in the range of 3.5%. We find a significant increase in these returns as workers age. From age 24 through age 31, the return to math skills increases by 13 percentage points. Most interestingly, we find that the returns to mathematical ability are highly non-linear. For instance, while moving from the $50^{th}$ percentile of the math test score distribution to the $85^{th}$ percentile results in a 30 percent wage increase, the same returns are observed by moving from the $95^{th}$ to the $99^{th}$ percentile. These patterns become starker for students in the top percentile, who outearn those in the next percentile by upwards of 14 percent and those in the $90^{th}$ percentile by upwards of 50 percent. We explore whether the same results are present for language test scores and find that the estimated returns are small and follow a linear pattern.

We find that educational attainment plays explains part of the return to skills, as students at the top of the math test distribution attain more years of education by age 29. Nonetheless, these students separate themselves along other measures of attainment. First, higher-skilled students have a higher likelihood of having completed a five-year or graduate degree by age 29. Furthermore, these students graduate with degrees from the highest-paying business, STEM or medicine majors (Hastings et al., 2014). Most remarkably, top-tier students differentiate themselves through the quality of the university attended. While 41% of $100^{th}$ math percentile students graduate from Chile's top two elite universities, only 13% of those in the $95^{th}$ percentile do so.

Building on the literature on the importance of firms in the labor market, we analyze the extent to which firm-worker matching mediates the returns to skills. We follow Haltiwanger et al. (2018) and present various definitions of firm quality. Across these measures, we find that high-skilled workers match with high-quality firms immediately upon labor market entry. While there is mobility up the job ladder for all workers across the test score distribution, assortative matching takes place early on in workers' careers. For instance, at age 25, an average worker in the top test score percentile works in a firm in the $80^{th}$ percentile of the firm-wage distribution, whereas one in the median of the math ability distribution works in a firm in the $56^{th}$ percentile. We estimate an augmented human capital equation and find that the estimated returns to math ability fall from 22% to 7-9% once detailed measures of education *and* firm quality are included. We find similar patterns for the non-linear estimates of the return to skills. To understand the importance of education and firm quality for the returns to skill, we conduct a decomposition analysis following Gelbach (2016) and find that the firm-matching component explains two-thirds of the change in the estimated returns to math ability. All in all, firms explain half of the aggregate returns to math test scores and these returns hold across the distribution.

The rest of the chapter proceeds as follows. Section 2 discusses the different administrative data sources used in the chapter. Section 3 estimates of the returns to skill, including linear and non-linear definitions of workers' skills. Section 4 characterizes heterogeneous educational attainment by skill levels and examines how such attainment mediates the returns estimated in the previous section. We additionally show matching patterns between high-ability workers and high-paying firms and again examine how worker-firm matching mediates the returns to abilities. We also present evidence from a Gelbach (2016) showing the share of the returns to skill explained separately by education and firm quality. Finally, in Section 5, we discuss the results and conclude.

### 3.2 Data Sources

We make use of three administrative data sources including information on students' mathematics and language test scores in a standardized $10^{th}$ grade, data on their high-school and post-secondary educational attainment and matched employee-employer data. We describe these data sources below.

**Test Score Information**

To measure students' pre-college skills, we take advantage of a standardized language and mathematics test score (SIMCE) administered to all enrolled $10^{th}$ graders in Chile. In particular, we focus on data from the 2001 and 2003 exams.[2] The goal of the SIMCE test is to evaluate the achievement of objectives and minimum content course requirements in Spanish language and mathematics among

[2]SIMCE was first implemented in 1998, but did not cover a nationally-representative sample. Starting in 2006, the exam has been carried out bi-annually and has recently changed to annual testing. SIMCE has algo begun testing students in $4^{th}$ and $8^{th}$ grade and it has included tests in English and science in recent years.

$10^{th}$ grade students. SIMCE scores are used by policymakers to monitor and evaluate schools by their performance in these two subjects and school-level scores are disseminated to parents to help them make their school choice decisions. On the other hand, individual scores are not given to students and are thus not used as a determinant of future educational attainment (or college entry), making them an attractive indicator of students' skills while in school.

The 2001 sample included upwards of 150,000 test-takers, whereas the 2003 sample exceeded 160,000. We exclude students who had either repeated a grade or missed one of the two exams, such that the main sample includes grade-for-age $10^{th}$ graders, covering the 1985 and 1987 birth cohorts, respectively. The full sample includes 131,200 students from the 2001 exam and 142,100 from the 2003 exam. For both test scores, the average score is 250 points with a standard deviation of 50 points. Within each test-taking sample, we create two measures of student ability. We first normalize students' math and reading scores imposing a normal distribution on the scores. For the second measure, we rank students by their percentile on the national test score distribution for each exam within the test-taking cohort. The large sample of test takers allows us to observe over 2,500 students in each percentile of the skill distribution, thus allowing us to estimate semi-parametric specifications of the returns to skill.

**Educational Attainment**

To determine students' educational attainment, we take advantage of two additional administrative data sources. We first link students in the SIMCE sample to data covering outcomes in high school for the universe of students enrolled in

Chilean high schools from 2002 through 2016. This link allows us to identify the time of high school graduation for students in the sample, and also identify high school dropouts: 12,300 students in the sample have not completed a high school degree by age 29.

We then take advantage of administrative records from the Higher Education Information System (SIES) for the 2005-2016 period. SIES is the governmental body within the Chilean Ministry of Education that manages and discloses official tertiary education statistics. This student-level data source tracks students higher-education enrollment and graduation patterns over time, indicating the institution of enrollment/graduation and the type of degree being being pursued/attained each year. As a result, we are able to identify the universe of higher education graduates in Chile and observe in detail the level of degree attained, the granting institution and the field of degree. We find that over 92,000 students in the sample had received a higher education degree by age 29, with 47% of them receiving a five-year bachelor's degree, 30% earning an associates degree, and the remaining 13% getting a four-year bachelor's degree.[3]

**Labor Market Outcomes**

To explore labor market outcomes, we use Unemployment Insurance (UI, *Seguro de Cesantia*) data, which contains matched employee-employer data for all formal sector employment contracts. The UI database has records of all formal workers' monthly earnings from November 2002 through June 2016, including up-

[3]In defining the years of education completed for each students, we include the years of enrollment for those who enrolled in higher-education but did not complete a degree by age 29. Nonetheless, we classify these students together with high-school graduates when defining educational attainment by highest degree completed.

wards of seven million workers. UI tracks workers with a unique identifier, and it includes information on workers' monthly earnings, sector of employment, state of residence, and observable characteristics, including gender and age. For every job held by every worker, we observe the month of entry and exit, allowing us to construct a measure of months worked in each year and a measure of total formal sector experience. The empirical analysis below includes workers' labor market outcomes when aged between 24 and 31. We impose this restriction to avoid including part-time employment for students still enrolled in college, but we plan on relaxing this restriction in the future. We note, however, that the measures of experience and tenure are constructed including workers' entire labor market trajectories.

UI also includes unique identifiers for each firm, allowing us to construct a longitudinal panel of the universe of firms in Chile and thus correctly workers' firm-tenure as they remain employed at the same firm. As UI covers the universe of formal sector firms in Chile each year, we use this information to construct an annual measure of firm quality, ranking firms by their average monthly wages and their median monthly wages, as in Haltiwanger et al. (2018). We combine the worker and firm panels to track worker flows across establishments.

## 3.3 Returns to Skills

The existing literature has largely estimated the returns to skill in linear specifications, but has sometimes included quadratic terms or interactions between different measures of abilities. We first estimate a linear model exploring the returns to $10^{th}$

grade test scores, $\boldsymbol{\theta_i}$:

$$lnwage_{it} = \beta_0 + \beta_1 \boldsymbol{\theta_i} + \beta_2 age_{it} + \lambda_t + \varepsilon_{it} \tag{3.1}$$

The first panel of Table 3.1 presents the estimated returns to SIMCE test scores in Chile. In the first column, we find that a one standard deviation increase in the math score is associated with a 22 percent increase in wages for workers aged 24-31. In the second column, we instead explore the returns to language scores, and find estimated returns in the range of 16 percent. In the last column, we find that, upon including both measures of test scores, the estimated returns to language skills fall significantly, down to 3.5 percent, whereas the returns to the math score remain large and significant, in excess of 20%. The returns to mathematical ability fall in line with the results presented by Deming (2017), who finds a return of 20.3 percent to cognitive ability, defined using AFQT scores, using NLSY data. However, these returns are significantly larger to those found by Lindqvist and Vestman (2011), who find a return of 8.6 percent to cognitive ablity in Sweden. We also analyze whether the returns to skill increase as workers age, following insights from Heckman et al. (2006) and MacLeod et al. (2017), who find increasing returns to schooling as workers accumulate labor market experience in the following specification:

$$lnwage_{it} = \beta_0 + \beta_1 \boldsymbol{\theta_i} + \beta_2 \boldsymbol{\theta_i} \times age_{it} + \beta_3 age_{it} + \lambda_t + \varepsilon_{it} \tag{3.2}$$

The second panel in Table 3.1 presents the estimated returns by age. The first column shows that the returns to math-related skills increase by 2.4 percentage points by year. These returns remain large and significant when we include language test scores in column 2. The returns to math test scores increase by 1.9 percentage points by year, whereas the returns to language skills increase by just 0.8 percentage points, though these returns are significant as well. As discussed above, estimating the returns to ability in a linear functional form may miss critical features of the returns to skills, such as allowing for differential returns across the distribution. Taking advantage of the large sample size of SIMCE test takers, we estimate the return to ability across each percentile of the distribution, $\boldsymbol{\theta_{ji}}$, in the following regression:

$$lnwage_{it} = \beta_0 + \sum_{j=1}^{100} \beta_j \boldsymbol{\theta_{ji}} + \beta_2 age_{it} + \lambda_t + \varepsilon_{it} \tag{3.3}$$

Figure 3.1 presents the estimated returns to math ability from equation (3.3).[4] The omitted category represents students in the bottom percentile of the math distribution, and the figure clearly shows the non-linearity in the returns to skill. On average, students in the median of the math score distribution outearns those at the bottom by 22 log points, whereas those in the $90^{th}$ percentile outearn those in the median by 36 log points. The figure clearly shows that the difference in returns clearly grows as we move up the skill distribution. For instance, while a student in the $90^{th}$ percentile of the math test score distribution outearns one at the $50^{th}$ percentile by 36 log points at age 30, the same returns are observed from moving

[4] We estimate equation (3.3) including one-hundred dummy variables for a student's ranking in each of the mathematics and language test scores.

from the $95^{th}$ through the $99^{th}$ percentile of the math test score distribution. In fact, the return to moving from the $99^{th}$ percentile of the test score distribution to the top centile results in a wage gain of 14 percent. On the other hand, an equivalent move up a math-skill percentile results in a wage gain of, at most, 3 percent at any percentile below the $90^{th}$. These results lend credence to the possibility that the market rewards skill "superstars", where students are the very top of the distribution significantly outearn those just a few percentiles below. In Appendix Table B.1.1, we present estimated returns from a semi-parametric regression where workers are placed in one of six bins of math achievement.[5] The results are similar to those presented above: top math achievers outearn those in the middle of the distribution as well as their peers just a few percentiles below them.

A potential concern with using students' ranking in the test score distribution as a measure of ability is that it may mask larger gaps in average test scores between the $100^{th}$ and $99^{th}$ percentile vis-a-vis the equivalent difference between the $96^{th}$ and $95^{th}$ percentile, thus creating mechanical non-linearities in the returns to skill.[6] To address this concern, we estimate rank-rank regressions analyzing the relationship between skill rankings and within-cohort wage rankings. This empirical approach has been extensively used in the intergenerational mobility literature (Chetty et al., 2014; Chetty and Hendren, 2018). We define the outcome variable as students' wage ranking relative to their birth cohort's peers' wages for each year they participate

[5]The six bins are: below median, between $p_{51}$ and $p_{80}$, the second top decile, between $p_{90}$ and $p_{95}$, between $p_{95}$ and $p_{99}$, and the top math percentile.

[6]In the future, we plan to explore alternative rankings of students' test scores, such as placing students in a bins comprising 0.1 standard deviations of each test score measure, thus ensuring the gap across test score percentiles remains equal across the distribution.

in the labor market. The semi-parametric rank-rank regression is then:

$$rank_{it}^{w} = \beta_0 + \beta_1 rank_i^{\theta} + \epsilon_{it} \tag{3.4}$$

where $rank_{it}^{w}$ denotes person $i$'s within-cohort wage ranking in year $t$. We present the estimated results in Figure 3.2 for workers at age 29. The first panel shows the relationship between ranking in the math test score and wages, and the non-linearities in the estimated returns are present as wells. We first note that workers at the median of the math skill distribution are, on average, in the $46^{th}$ percentile of their cohort's wage distribution at age 29, outpacing their counterparts at the bottom of the distribution by 13 percentiles. Similarly, there is a significant increase in the average rank for workers in the $90^{th}$ percentile of the math distribution, whose wages are, on average, at the $63^{rd}$ percentile by age 29. Nonetheless, these differences become larger when focusing on those at the very top of the math distribution: while $95^{th}$ percentile math-achievers reach the $68^{th}$ percentile of the wage distribution, those in the top math percentile far out-rank them by reaching, on average, the $78^{th}$ percentile of the wage distribution. These results confirm the previously-estimated returns to being a math skill 'superstar.'

Panel B shows results from the same regression, but we instead focus on the return to students' ranking in the language exam. The results are strikingly different. While students in the median of the language distribution fall on the $41^{st}$ percentile of the wage distribution, those in the $90^{th}$ language percentile are, on average, in the same percentile of the distribution. There is a slight difference for students in the top language centile, moving up two percentiles in the wage distribution, yet this difference is significantly smaller than the corresponding result for math ability. These findings confirm those presented in Table 3.1 and Figure 3.1, showing both large (and non-linear) returns to math ability accompanied by small (and largely linear) returns to language skills. In the next section, we explore how different measures of educational attainment and firm quality vary by students' ranking in the math test score distribution.

## 3.4 Mechanisms

### 3.4.1 Educational Attainment

Education can mediate the returns to skill through either signaling or human capital mechanisms (or both): in a signaling world, it is less costly for skilled students to attain more years of education, whereas in the human capital story, skilled students are capable of learning more while in school, thus become more productive and earning higher wages. The existing literature has confirmed the mediating role of education. Lindqvist and Vestman (2011) find that the estimated returns to cognitive ability in Sweden fall from 0.086 to 0.050 once education is controlled for. Similarly, Deming (2017) finds that the returns to AFQT in the NLSY fall from 0.203 to 0.129 upon controlling for years of education attuned. Prada (2014) estimates a discrete choice model and reaches a similar conclusion, finding that one-third of the return to cognitive ability is explained by educational attainment.

In the first panel of Figure 3.3, we show the average years of education completed by students' ranking on the math SIMCE through age 29. Unsurprisingly, we find that higher skilled students tend to achieve more years of education. For instance, students below the $20^{th}$ math percentile finish about less than or just twelve years of education, whereas those in the $80^{th}$ percentile attain an additional two years by age 29. While moving to the top of the math distribution results in an additional 1.5 years of completed education (reaching upwards of 15.5 total years), the stark non-linearities present in Figures 3.1 and 3.3 are not observed in this panel. As Rodriguez et al. (2016) had previously found significant heterogeneity in the returns to different types of degrees in Chile, such that five-year bachelor's students outearn those with a four-year degree and associate's recipients, in Panel B, we explore the relationship between math skill ranking and the type of degree attained by 29. We focus on the share of students who have either received a five-year bachelor's degree or a graduate degree by 29. The non-linear patterns in educational attainment by mathematical skill begin to emerge more clearly. As a result, while just 13 percent of students in the median of the math distribution attain either of these degrees by age 29, 45 percent of students in the $90^{th}$ math percentile do so by age 29. As with wages, there is a significant difference within students in the top decile, such that over 70 percent of students in the top of the math distribution earn one of these degrees, far outpacing the rest of the top math test score decile. These results indicate that high skilled students differentiate themselves from their lower-skilled counterparts on dimensions of educational attainment beyond years completed.

In Figure 3.4, we further explore sorting along other educational dimensions. In the first panel, we examine the share of students receiving either a five-year bachelor's degree or a graduate degree in either health, science, or medicine. We focus particularly on these degrees since Hastings et al. (2014) have shown these degrees deliver the largest returns in Chile. Again, we find significant differences across the math skill distribution. Whereas 57 percent of students in the top percentile attain one of these degrees by age 29, this share drops by half, down to 28 percent, for students in the $90^{th}$ percentile. At the same time, just 5 percent of students in the median of the distribution attain either of these degrees. In the second panel, we analyze the share of students who receive either a 5-year or a graduate degree from one of Chile's two elite universities, *Universidad Catolica* or *Universidad de Chile*. These two universities are of particular interest, as Zimmerman (2017) has found that admission to either one results in a sizable increase in the probability of reaching the top 0.1% of the income distribution. The non-linearities become even starker in this context: 41% of top achieving math students attain a degree from these universities, compared to just 0.2 percent for students in the median of the distribution. Similarly, students at the top are nine times as likely as their counterparts in the $90^{th}$ percentile to receive a degree from an elite institution.

To formally explore how education mediates the returns to skill, we estimate a human capital regression, including measures of labor market experience and firm-tenure, as in Altonji and Williams (2005) and Topel (1991):

$$lnwage_{it} = \beta_0 + \beta_1 \boldsymbol{\theta_i} + \beta_2 age_{it} + \beta_3 \boldsymbol{S_{it}} + \gamma_1 f(exper_{it}) + \eta_1 tenure_{it} + \lambda_t + \varepsilon_{it} \quad (3.5)$$

$\boldsymbol{S_{it}}$ represents different definitions of educational attainment. We present the results in the first panel of Table 3.2. The first column replicates the results from the first table, showing returns to math skills of upwards of 20 percent. As we move across columns, the inclusion of labor market experience and tenure increase the estimated returns to ability, given that low-ability students enter the workforce earlier and have thus acquired more experience through their late twenties. The last column includes the years of education completed by worker $i$ in year $t$.[7] These results fit in with previous findings from the existing literature, as the estimated returns to math skills fall from 21.7 percent to 16.2 percent, and the returns to language fall from 4.8 percent to 1.4 percent. In Panel B, we explore how the estimated returns change as we include different definitions of educational attainment. Including dummy variables for the types of degrees attained in column (3) does not have a significant effect on the estimated returns to skill. However, when we define attainment by the interaction of degrees and university quality, as in column (4), or by the interaction of types and fields of degrees (column 5), the estimated returns fall to 15 and 14 percent, respectively. Table 3.2 confirms that education mediates the returns to skills, but we argue that different definitions of attainment explain varying shares of these returns. In Appendix Table B.1.2, we estimate a semi-parametric version of equation (3.5). We also find that education mediates non-linear returns to math ability, as the estimated returns for top percentile students, relative to those in the second decile, fall from 62 percent in the baseline regression (column 1) to 42 percent once fields and types of degrees are taken into account (column 5). We next explore

[7]While in Figures 3.3 and 3.4, we had explored how educational attainment at age 29 varied with workers pre-college skills, the measure of education in Table 3.2 defines attainment as the concurrent number of years completed by worker $i$ by year $t$ instead on focusing on their completed educational history.

the relationship between skills and firm quality.

### 3.4.2 Firm Quality

An extensive literature in the spirit of Abowd et al. (1999) (AKM) has found that firms explain an important share of the variance of wages. For instance, Card et al. (2013) find that firm quality explains upwards of 20 percent of the variance of log wages in West Germany. As noted above, empirical models following AKM can be used to examine patterns of assortative matching between high-quality workers and high-quality firms, and Card et al. (2013, 2018) have found evidence of positive assortative matching. However, this approach identifies worker quality through labor market outcomes, which does not offer a direct mapping to workers' pre-labor market characteristics, such as ability. In this section, we take advantage of a well defined measure of worker ability, represented by SIMCE scores, to document patterns of assortative matching in the labor market and explore how they firms mediate the returns to skill.

We follow Haltiwanger et al. (2018) and create an annual ranking of firm quality based on employment-weighted measures of: (1) average wages, (2) median wages, (3) $p_{25}$ wages.[8] we also explore how the worker-firm match evolves as workers acquire labor market experience, following insights from Haltiwanger et al. (2018), among others, who explore the types of workers who move up the firm-quality ladder. We first examine how workers match with firms across the ability distribution in

[8]We have also estimated a measure of firm quality by using the estimated firm fixed-effect from an AKM regression for all workers in the UI data. However, Haltiwanger et al. (2018) note that these estimates of firm quality follow only if a strict exogenous worker mobility assumption is met. The results, available upon request, are similar with either measure of firm quality.

the following specification:

$$\boldsymbol{\varphi_{ift}} = \beta_0 + \beta_1 \boldsymbol{\theta_i} + \beta_2 age_{it} + +\lambda_t + \varepsilon_{it} \tag{3.6}$$

In equation (3.6), $\boldsymbol{\varphi_{ift}}$ represents one of the above-mentioned measures of firm quality for worker $i$ in year $t$. Figure 3.5 shows the relationship between math skills and firm ranking, measured by the firm's average monthly wage. The first panel shows evidence of positive assortative matching between high-skilled workers and high-paying firms as early as age 25. We note that while young workers across the ability distribution tend to work in firms with average wages above the median, there is significant sorting as we move up the distribution. As a result, an average worker at the median of the math test score distribution works in a firm in the $56^{th}$ percentile of the firm distribution, whereas one in the $90^{th}$ skill percentile is on average thirteen ranks higher in the job ladder. Furthermore, similar to the results shown for educational attainment and wages, the top test score achievers are employed, on average, at $80^{th}$ percentile firms, far outpacing their counterparts ten math skill percentiles below.

The second panel presents evidence from the same regression, but for workers at age 29, where we find that high-skill workers are employed in higher-paying firms. We complement these results with alternative measures of firm quality in Appendix Figure 13, defining quality by firms belonging to the top 10% and top 5% of the wage distribution. The results again indicate a strong matching component, as, for instance, 53% of workers in the top math test score centile reach a top 10% firm, relative to just 31% of those in the $90^{th}$ percentile. These patterns become starker as we focus on firms in the top 5 percent of the wage distributions.

The results presented in Figure 3.5 show clear movement up the job ladder for workers across the test score distribution. Interestingly enough, low ability workers seem to be moving further up the distribution vis-a-vis their high ability counterparts. For instance, those in the median of the math skill distribution have gone up, on average, from a firm in the $56^{th}$ percentile of the wage distribution at age 25 to one in the $63^{rd}$ percentile at age 29. Meanwhile, those in the top percentile of math test scores have moved up, on average, four ranks. These results fit in with Haltiwanger et al. (2018), who had previously found that low educated workers were most likely to move up the job ladder. In fact, the evidence presented here offers a potential explanation behind their results: as high ability workers match with high quality firms early on their labor market career, there does not seem to be much room for further moves up the ladder as they age.

Given the previously-found importance of firms for explaining wage inequality and the fact that there is positive assortative matching between workers and firms, we examine how firms mediate the returns to skill in the following specification:

$$lnwage_{ift} = \beta_0 + \beta_1\boldsymbol{\theta_i} + \beta_2 age_{it} + \beta_3\boldsymbol{S_{it}} + \beta_4\boldsymbol{\varphi_{ft}} + \gamma_1 exp_{it} + \eta_1 ten_{it} + \lambda_t + \varepsilon_{ift} \quad (3.7)$$

Table 3.3 presents the estimates from equation (3.7). The first column presents the returns to math and language test scores from the human capital equation with education defined by the fields of degree attained.[9] In the second column, we include a linear term of worker $i$'s firm of employment's average wage ranking in year $t$ and the estimated returns to math test scores fall from 14.1 percent to 8.1 percent. In the third column, we use the same measure of firm quality, but include one hundred dummy variables for each of the firm quality percentiles. In this specification, the estimated returns to math skills fall further to 7.6 percent, or almost half of the baseline returns in the human capital equation. Finally, the last two columns show that defining firm quality by the median and the $p_{25}$ wage ranking still lower the estimated returns to math skills, though the change is smaller in magnitude than for the preferred definition.

In Figure 3.6, we present the estimates from a semi-parametric estimation of equation (3.7), having included dummy variables for each percentile for the math and language test scores, as in equation (3.3). We find similar results as in the linear specification. The estimated returns to being in the top math percentile vis-a-vis those in the bottom of the distribution falls from 120 percent in the baseline specification to less than 50 percent once field of degree dummies and firm quality dummies are included in the regression. Similarly, the estimated return to workers in the $90^{th}$ percentile of the math distribution compared to those at the median falls from 46 percent to 19 percent upon including educational and firm quality. The results presented in this sub-section indicate that both components play a critical role in mediating the returns to ability. Nonetheless, the results presented so far

[9] The results are similar when defining attainment by a measure of university quality. These results are available upon request.

do not allow us to estimate the separate contribution of the education and the firm channel, as any such estimation would suffer from sequence dependence, where the estimated contribution of each component would change by the order in which controls are added to the model. In the next section, we present evidence from a decomposition proposed by Gelbach (2016) to estimate how education and firm quality separately mediate the returns to skills in the labor market.

### 3.4.3 Gelbach Decomposition

We have so far shown that higher-skilled individuals are morel likely to attain more years of education, graduate from higher-quality universities and higher paying fields, as well as work in high-paying firms. In order to explore how these both education and firm quality mediate the returns to skill, it is useful to re-write equations (3.1) and (3.7) as follows:

$$lnwage_{ift} = \beta_0^B + \beta_1^B \boldsymbol{\theta_i} + \beta_2^B age_{it} + \gamma_1^B exp_{it} + \eta_1^B ten_{it} + \varepsilon_{ift} \tag{3.8}$$

$$lnwage_{ift} = \beta_0^F + \beta_1^F \boldsymbol{\theta_i} + \beta_2^F age_{it} + \gamma_1^F exp_{it} + \eta_1^F ten_{it} + \beta_3^F \boldsymbol{S_{it}} + \beta_4^F \boldsymbol{\varphi_{ft}} + \varepsilon_{ift} \tag{3.9}$$

$\beta_1^B$ provides an estimate of the net returns to pre-college skills in the labor market, whereas $\beta_1^F$ represents the estimates after having accounted for educational attainment and firm matching decisions. To understand the separate contribution of these two factors, we follow Gelbach (2016) who presents a methodology based on the omitted variable bias formula. In this analysis, the econometrician can recover the separate effect of any omitted variables from the estimation of equation (3.9) by following these steps. First, estimating the full model (equation (3.9)) and recovering ($\beta_1^F,\ \beta_2^F,\ \gamma_1^F,\ \eta_1^F,\ \beta_3^F,\ and\ \beta_4^F$). Second, running a regression of $\boldsymbol{S_{it}}$ on $\boldsymbol{\theta_i}$, yielding a coefficient $\tau_S$, and one for $\boldsymbol{\varphi_{ft}}$ on $\boldsymbol{\theta_i}$, yielding $\tau_\varphi$. The two $\tau$ coefficients are the differences in educational attainment and firm quality by math ability, such that the difference in the changes in the estimated returns to ability is given by: $\beta_1^F - \beta_1^B = \beta_3^F \tau_S + \beta_4 \tau_\varphi$. The share of change the in $\beta_1$ explained by educational attainment is given by $\beta_3^F \tau_S$, and the share explained by firm matching is $\beta_4^F \tau_\varphi$.

We follow this procedure and present the estimates from the linear returns to ability in Table 3.4. As seen in Table 3.3, upon including dummy variables for the field of degree attained and for the firm ranking percentiles, the estimated returns to math skill fall from 21.7 percent to 7.6 percent. We find that differences in educational attainment across the skill distribution explain less than one-third of the estimated change in the returns to skill, with worker-firm assortative matching accounting for almost 70 percent of the change. The returns to language skills similarly fall from 4.8 percent to 0.2 percent, and again, this change is largely driven by the worker-firm match component. In Figure 3.7, we present the results from this decomposition for analyzing the sources behind the changing non-linear returns to skill discussed in Figure 3.6. The baseline returns to being in the top math percentile fall from 120 percent to 102 percent (relative to the bottom math students) once fields of degrees are accounted for, yet drop significantly, down to 50 percent, due to the worker-firm match component. A similar result is found for the returns to workers in the median of the skill distribution: over two-thirds of the non-linear returns to skills are explained by high ability workers working in high-quality firms. While the results may seem surprising given the strong relationship between pre-college skills and various measures of educational attainment, these findings fit in with the above-mentioned literature which finds that firms play a critical role in determining labor market outcomes. While future work should consider potential interactions between education firm matches, these results results indicate that good workers and good firms meet early and often in the labor market, and that this mechanism explains a sizable share of the returns to skill.

## 3.5 Conclusion

The technological revolution which has taken place in recent decades across both developed and developing countries has brought increased attention to understanding the returns to pre-labor market abilities. While an extensive literature has found sizable returns to skill, these papers have largely relied on parametric functional forms, often focusing on the linear returns to skill. In this chapter, we have shown that this literature has so far missed a critical component of the returns to skill, by failing to consider the existence of skill 'superstars' and exploring the returns for these workers. To better understand the channels through which skilled workers earn higher wages, we have also explored different measures of educational attainment through which these workers may differentiate themselves from their lower-skilled peers. While the existing literature has often documented that high skilled workers attain more years of education, we have shown that this relationship is significantly stronger when defining attainment by university and degree quality. Furthermore, this distinction is critical when defining ability in a non-linear functional form.

Lastly, building off an extensive literature showing that firms play a large role in determining wage dispersion, we have thoroughly examined the importance of firms for understanding the returns to skills. In a first in the literature, we have shown that high-ability workers match with high-paying firms immediately upon labor market entry and that there is subsequent movement up the job ladder for young workers. In fact, a decomposition analysis indicates that firms explain at least two-thirds of the explained component of both the linear and non-linear returns to skills. Further work is required on this topic, in particular in exploring interactions between educational attainment and firm quality, but we have shown that the interaction between workers' abilities and firms is critical for understanding labor market outcomes. In the next chapter, I extend this analysis by considering how early life skills affect individuals' human capital decisions, focusing in particular on their progression through higher education and college major choices. By considering the returns to college majors, the next chapter seeks to decompose the causal returns to human capital investment decisions, thus building on the contributions in this chapter.

## 3.6 Tables and Figures

**Table 3.1:** Returns to Ability in Chile (Monthly Wages)
**Panel A.** Linear Returns to Ability

| | (1) | (2) | (3) |
|---|---|---|---|
| Math | 0.224*** | | 0.203*** |
| | (0.0002) | | (0.0003) |
| Language | | 0.163*** | 0.034*** |
| | | (0.0002) | (0.0003) |
| Age | 0.054*** | 0.052*** | 0.055*** |
| | (0.0006) | (0.0002) | (0.0002) |
| Year FE | X | X | X |
| $R^2$ | 0.124 | 0.090 | 0.125 |
| Observations | | 10,170,432 | |
| Individual Observations | | 243,267 | |

**Panel B.** Linear Returns to Ability by Age

| | (1) | (2) |
|---|---|---|
| Math | 0.147*** | 0.141*** |
| | (0.0005) | (0.0006) |
| Math × Age | 0.024*** | 0.019*** |
| | (0.0001) | (0.0001) |
| Language | | 0.009*** |
| | | (0.0006) |
| Language × Age | | 0.008*** |
| | | (0.0001) |
| Year FE | X | X |
| $R^2$ | 0.123 | 0.124 |
| Observations | 10,170,432 | |
| Individual Observations | 243,267 | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples. SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized. Wages are measured monthly in 2010 Real CLP in the highest paid job.

**Table 3.2:** Human Capital Equation
**Panel A.** Educational Attainment

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Math | 0.203*** | 0.221*** | 0.217*** | 0.162*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Language | 0.034*** | 0.052*** | 0.0481*** | 0.0144*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| *Experience* | | 0.094*** | 0.0732*** | 0.110*** |
| | | (0.0003) | (0.0003) | (0.0003) |
| $Experience^2$ | | -0.004*** | -0.00426*** | -0.00503*** |
| | | (0.0003) | (0.0003) | (0.0003) |
| Tenure | | | 0.053*** | 0.047*** |
| | | | (0.0003) | (0.0003) |
| Year FE | X | X | X | X |
| Educational Attainment | | | | X |
| $R^2$ | 0.125 | 0.148 | 0.161 | 0.230 |
| Observations | | 10,170,432 | | |
| Individual Observations | | 243,267 | | |

**Panel B.** Definitions of Educational Attainment

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Math | 0.217*** | 0.162*** | 0.159*** | 0.152*** | 0.141*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Language | 0.048*** | 0.014*** | 0.014*** | 0.013*** | 0.021*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| *Experience* | 0.073*** | 0.110*** | 0.112*** | 0.112*** | 0.111*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| $Experience^2$ | -0.004 | -0.005 | -0.005 | -0.005 | -0.005 |
| | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) |
| Tenure | 0.053*** | 0.047*** | 0.047*** | 0.047*** | 0.047*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Year FE | X | X | X | X | X |
| Years of Ed. | | X | | | |
| Degrees Received | | | X | | |
| University Quality | | | | X | |
| Field of Degree | | | | | X |
| $R^2$ | 0.161 | 0.206 | 0.209 | 0.213 | 0.223 |
| Observations | | | 10,170,432 | | |
| Individual Observations | | | 243,267 | | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples.

SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized.

**Table 3.3:** Returns to Ability: Firm Quality Definition

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Math | 0.141*** | 0.081*** | 0.076*** | 0.099*** | 0.098*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Language | 0.021*** | 0.002 | 0.002 | 0.010*** | 0.016*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| *Experience* | 0.111*** | 0.067*** | 0.066*** | 0.064*** | 0.075*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| $Experience^2$ | -0.005 | -0.004 | -0.003 | -0.003 | -0.003 |
| | (0.00003) | (0.00003) | (0.00003) | (0.00003) | (0.00003) |
| Tenure | 0.047*** | 0.033*** | 0.033*** | 0.032*** | 0.032*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Year FE | X | X | X | X | X |
| Field of Degree | X | X | X | X | X |
| Firm Rank (Mean Wage) | | X | | | |
| Firm Rank Dummies | | | X | | |
| Firm Rank ($p_{50}$) | | | | X | |
| Firm Rank ($p_{25}$) | | | | | X |
| $R^2$ | 0.244 | 0.468 | 0.482 | 0.425 | 0.402 |
| Observations | | | 10,170,432 | | |
| Individual Observations | | | 243,267 | | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples. SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized. Wages are measured monthly in 2010 Real CLP in the highest paid job.

**Table 3.4:** Returns to Ability: Gelbach Decomposition

| | (1) Baseline | (2) Full | (3) Education | (4) Firm Quality |
|---|---|---|---|---|
| Math | 0.217*** | 0.076*** | 0.045*** | 0.096**** |
| | (0.0003) | (0.0003) | (0.0001) | (0.0002) |
| | | | [0.319] | [0.681] |
| Language | 0.048*** | 0.002*** | 0.017*** | 0.029*** |
| | (0.0003) | (0.0003) | (0.0001) | (0.0002) |
| | | | [0.369] | [0.631] |
| Year FE | X | X | | |
| Field of Degree | | X | | |
| Firm Rank Dummies | | X | | |
| $R^2$ | 0.161 | 0.468 | | |
| Observations | 10,170,432 | | | |
| Individual Observations | 243,267 | | | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples. SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized. Wages are measured monthly in 2010 Real CLP in the highest paid job.

**Figure 3.1:** Estimated Returns to Math Ability

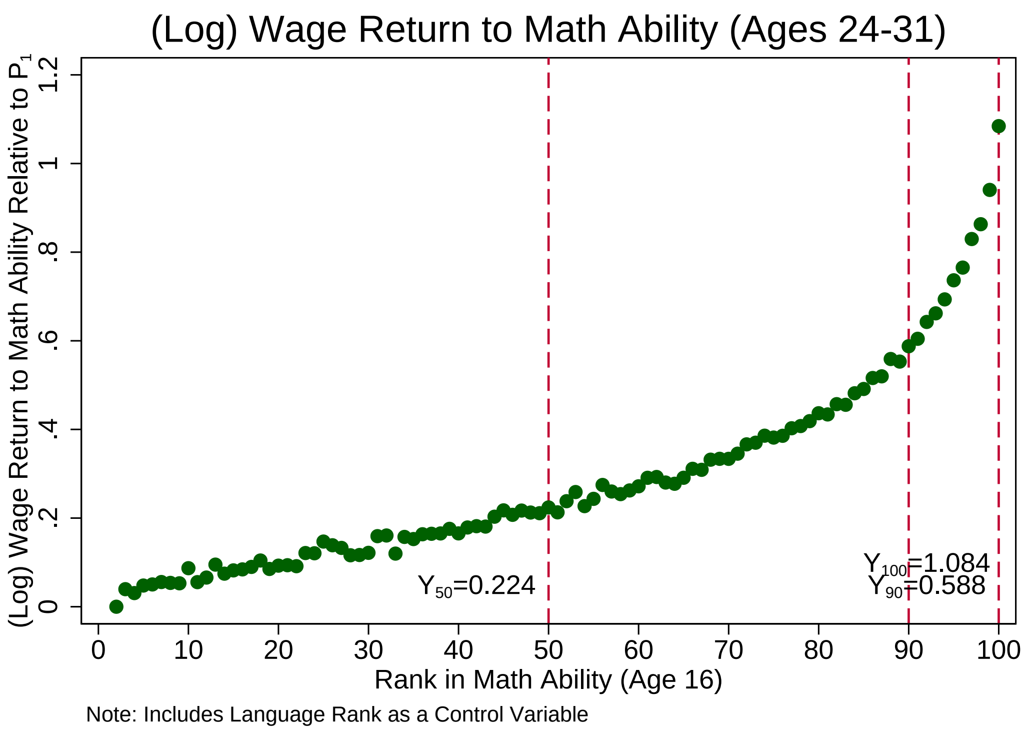
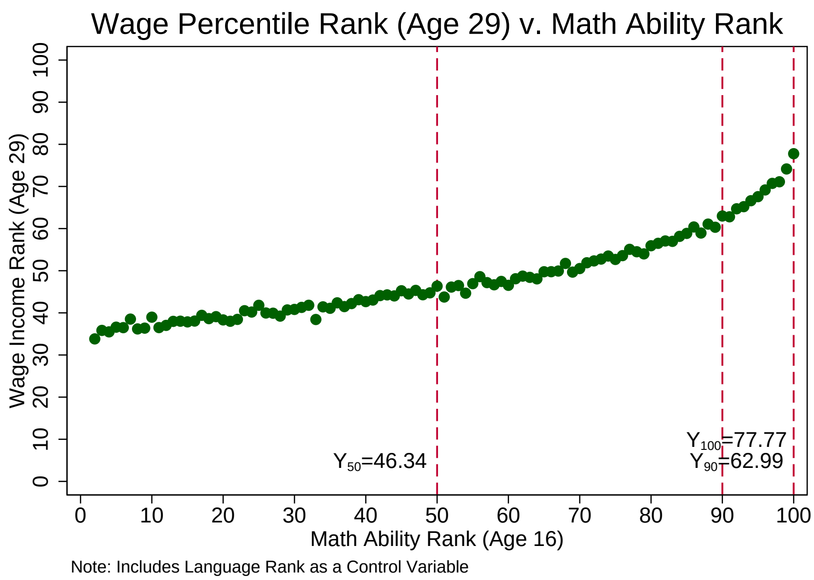


**Figure 3.2:** Rank-Rank Regressions



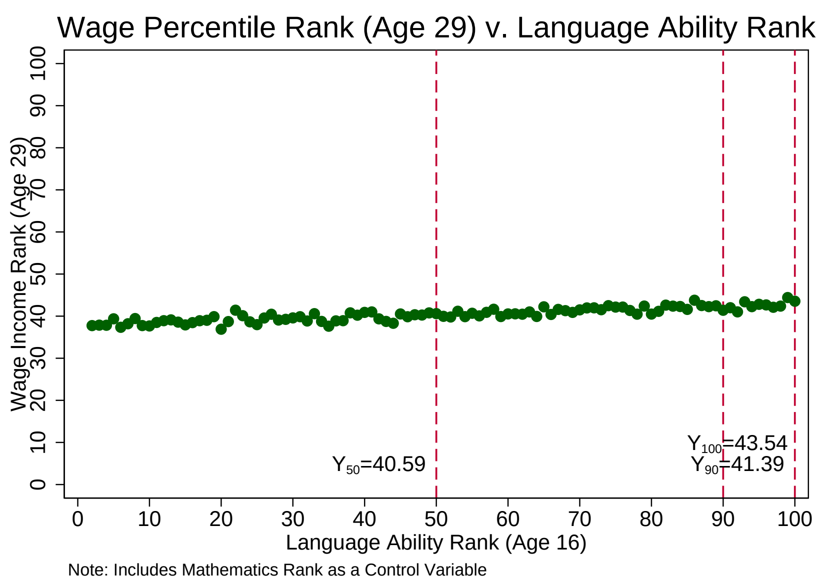
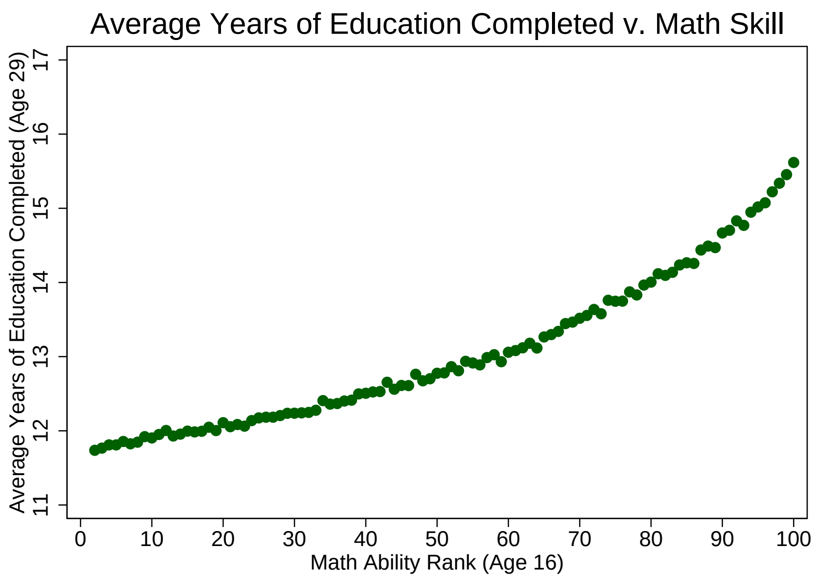



**Figure 3.3:** Educational Attainment



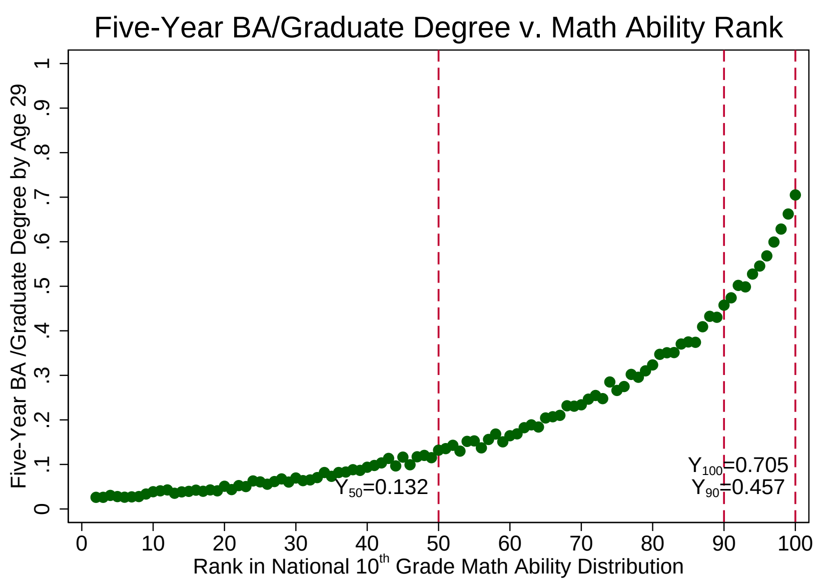



**Figure 3.4:** Educational Attainment: College Quality and Major

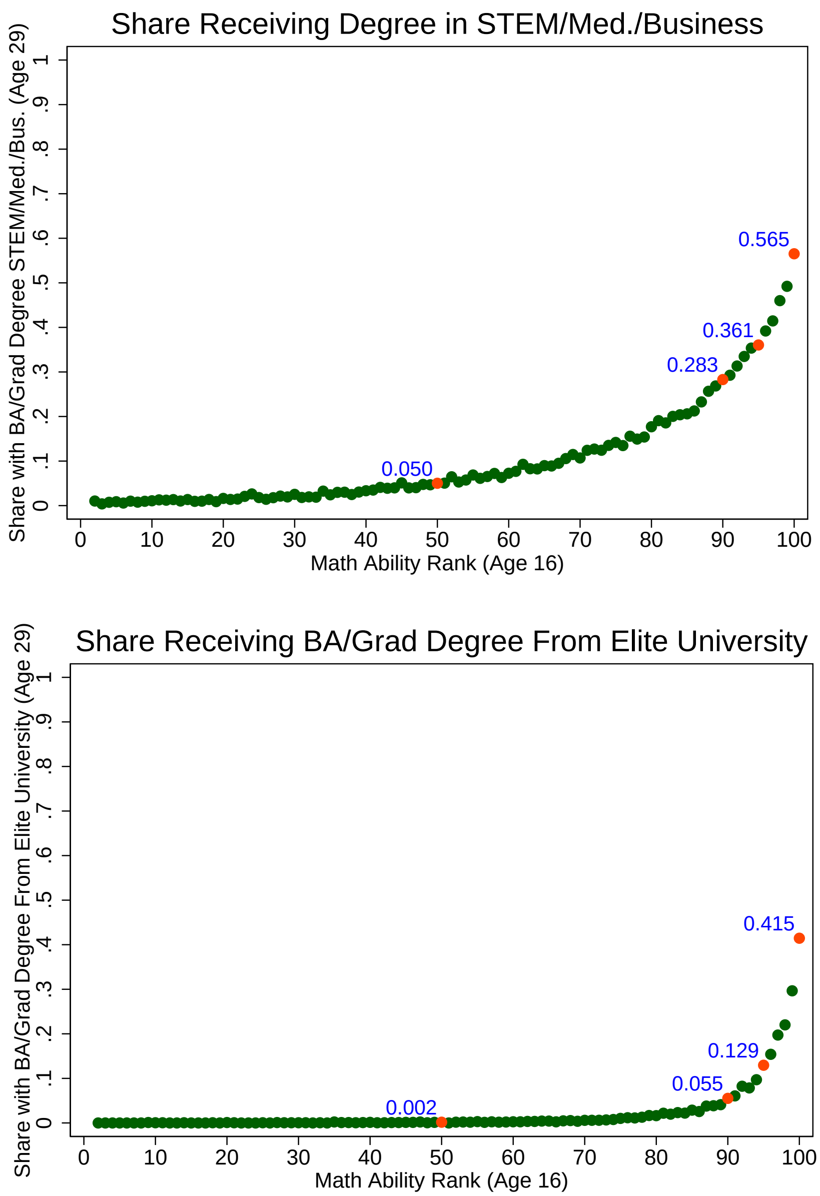
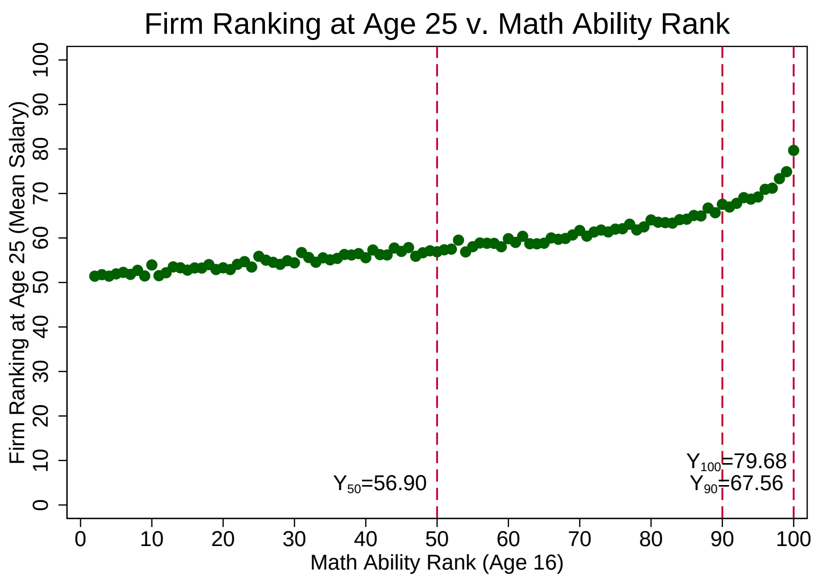


**Figure 3.5:** Assortative Matching: Firm Quality At Ages 25 and 29



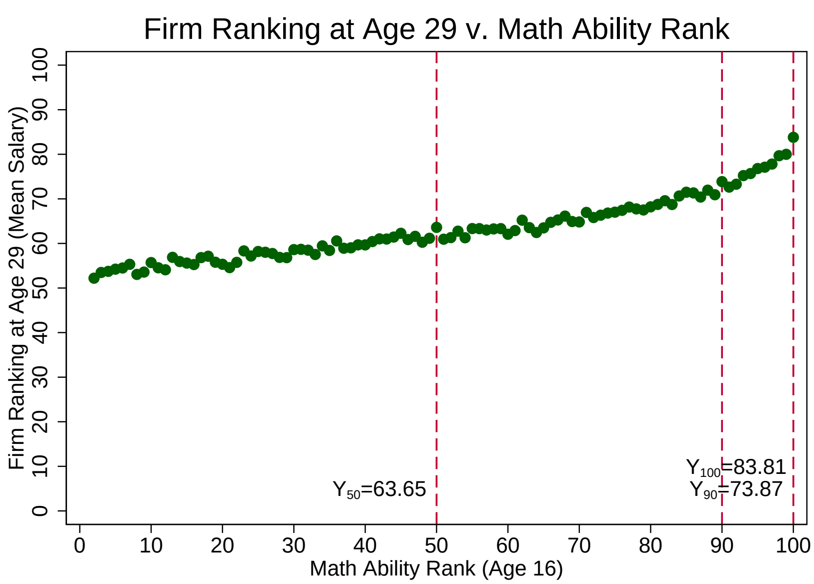



**Figure 3.6:** Human Capital Equation

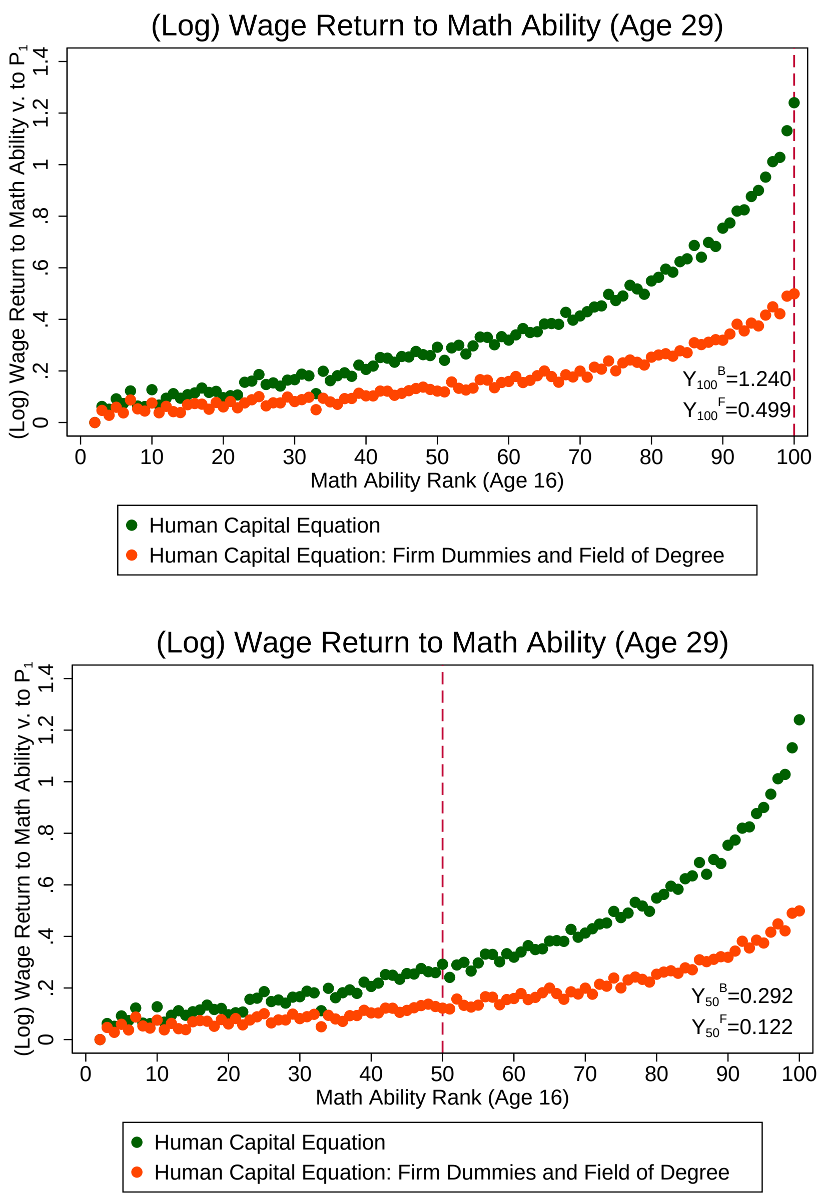
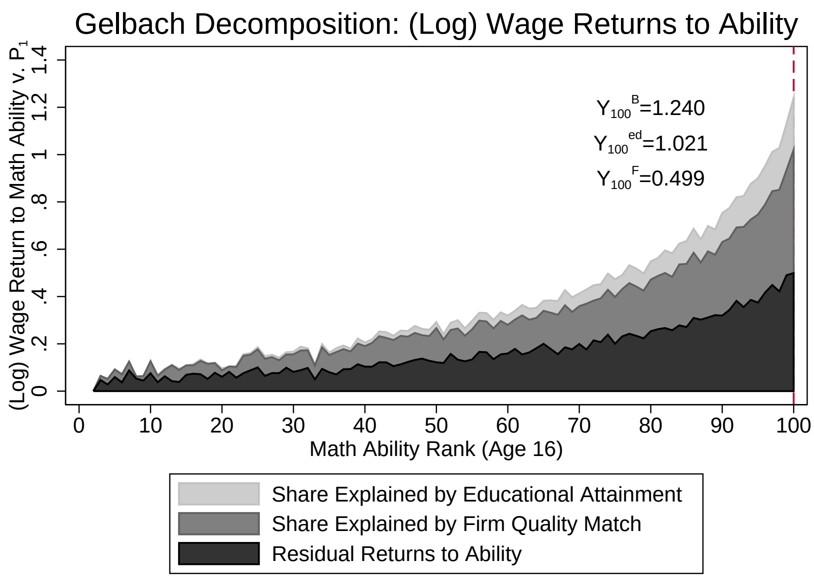
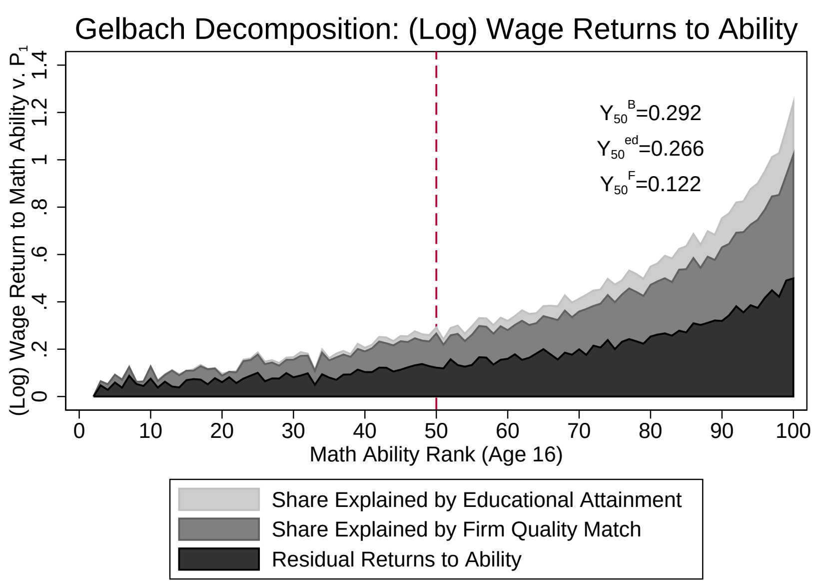


**Figure 3.7:** Gelbach Decomposition



# Chapter 4: What's Math Got to Do With It? Multidimensional Ability and the Gender Gap in STEM

## 4.1 Introduction

Women make-up just one fourth of recent graduates in math-intensive STEM majors in the United States (Kahn and Ginther, 2017). As these majors are among the highest-paying degrees, understanding the factors contributing to STEM participation gaps may offer guidance for narrowing gender gaps in labor market outcomes. In this context, colleges across the country have begun implementing policies aimed at boosting women's STEM enrollment rates (EOP, 2014). Nonetheless, while promoting enrollment in STEM majors is a critical first step for reducing gender gaps, half of initial enrollees fail to complete a STEM degree (Altonji et al., 2016), and the dropout rate is larger for women than for men (Kugler et al., 2017). As a result, understanding the factors which drive students to sort into majors and subsequently finish them can help in designing more effective policies aimed at promoting STEM participation and persistence.

In this chapter, I examine the interaction between pre-college math ability and major choices, focusing on women's enrollment and graduation from math-intensive STEM majors. Previous work has analyzed whether gender gaps in math test scores can explain the difference in STEM participation (Turner and Bowen, 1999; Xie and Shauman, 2003; Dickson, 2010; Riegle-Crumb et al., 2012; Justman and Méndez, 2018), yet test scores are affected by background characteristics and contaminated with measurement error, thus potentially mismeasuring the importance of math ability in gender STEM gaps.[1] Moreover, other skill dimensions, such as non-cognitive skills, may play an important role in determining students' college major choices as well as their STEM participation. In this context, I focus on the role of mathematical self-efficacy, which measures an individual's perceived ability to perform math-related tasks, in explaining gender gaps in math-intensive majors.

To understand students' enrollment and completion patterns given their pre-college ability, I present and estimate a sequential model of college progression. In this model, which builds on Heckman et al. (2016), Heckman et al. (2018), Humphries et al. (2017) and Rodrıguez et al. (2017), students first select a college major among five broad fields. In the second stage, they either complete their initial major, switch fields, or dropout of college altogether. In the last decision node, students are able to complete a graduate degree, or enter the labor market and earn hourly wages. At each stage, individual decisions and labor market outcomes are a function of observed characteristics and latent math and reading ability. I implement the model using Educational Longitudinal Study of 2002 (ELS) data, which follows a nationally-representative cohort of $10^{th}$ graders through age 26. ELS

[1] See Carneiro et al. (2003), Cunha et al. (2006), Heckman et al. (2006), Heckman et al. (2016), Borghans et al. (2008), and Prada and Urzúa (2017), among others. Niederle and Vesterlund (2010) have separately shown that competitive pressures explain part of the gender gap in math test scores.

data includes detailed information on multiple measures of math test scores, math class GPA, math self-efficacy measures, detailed information college major choices and early-career labor market outcomes.

I follow latent factor models to identify the distribution of unobserved ability through a measurement system. This approach allows me to correct for measurement error in test scores while controlling for the contribution of background characteristics to test scores. I take advantage of the various observed measures in the data to identify a non-cognitive skill component, math self-efficacy, along with a math problem solving factor and a reading ability component. I allow for these components to be correlated, relaxing the factor independence assumption imposed in previous work and fitting in with the recent literature on latent factors (Prada and Urzúa, 2017). Furthermore, as I estimate the model separately by gender, I can examine whether gender gaps in math test scores overstate those in latent math ability and whether the correlations across the latent ability components differ between men and women.[2] I first find that the gender gap in latent math problem solving ability is 40 percent smaller than the 0.30 standard deviation gap in math test scores. This result follows from the finding that math-course GPA reflects problem solving ability and women outperform their male peers in this dimension. The problem solving component is highly correlated with the math self-efficacy component, though the correlation is lower for women. There is thus a relative 'lack' of high-performing

[2] I interpret the components of latent ability to be fixed by the time of initial major choices, but do not consider these measures to be constant from birth. This assumption follows from an extensive literature showing gender gaps in math performance expand in elementary and middle school (Kahn and Ginther, 2017). Huang (2013) shows a similar pattern in math self-efficacy between middle- and high-school.

women who are confident in their math abilities vis-a-vis their male counterparts.

I find that math problem solving ability and self-efficacy are strong predictors of STEM enrollment for both men and women, and this decision is non-linear, as an increase in self-efficacy at the top of the problem solving distribution has a larger impact on enrollment than one for low math achievers. For instance, for women in the top math problem solving decile, only 2% of those who are in the bottom self-efficacy decile enroll in STEM, whereas 13% of those in the top decile do. As a result, the relative lack of women at the top of the joint skill distribution reduces their participation in math-intensive majors. On the other hand, gender differences in latent problem solving ability explain less than ten percent of the gap in STEM enrollment, fitting in with recent work highlighting the importance of preferences in driving STEM gaps (Zafar, 2013; Wiswall and Zafar, 2014, 2017). At the same time, as math self-efficacy explains an additional 7 percent of the enrollment gap, I remark the importance of considering multiple dimensions of ability when considering STEM participation decisions.

In terms of subsequent STEM completion, while 60% of men initially enrolled in these majors end up graduating, fewer than 45% of women do so. There is re-sorting on the problem solving component for both men and women, such that only the highest math-achievers graduate from these majors. However, self-efficacy plays a far larger role for women than it does for men in leading to degree completion. 55% of male enrollees in the bottom self-efficacy quintile complete a STEM degree, rising slightly to 64% for those in the top quintile. On the other hand, while only 22% of female STEM enrollees in the bottom self-efficacy quintile successfully complete a degree, the completion rate for those in the top quintile is almost three times as large, reaching 64%. As a result, a shortfall in this non-cognitive negatively affects women's STEM participation both at enrollment and graduation and accounts for 20% of the gender gap in STEM completion rates among STEM enrollees.

Despite the efforts aimed at increasing women's STEM participation, Altonji et al. (2012, 2016) note there is limited causal evidence on gender-specific returns to college majors. Estimating these returns is critical for understanding whether STEM participation would in fact improve women's labor market outcomes. I note that in the context of policies aimed at increased STEM enrollment, policymakers should be interested in the returns to enrollment, rather than on the returns to graduation. These parameters are different, as the former allows for the possibility that students may fail to complete their initial major, whereas the latter assumes successful completion. Moreover, since specific components of the ability vector may have differential effects across college majors, the returns to STEM majors may not be uniform for all students. In this context, estimating heterogeneous returns to majors allows for the correct identification of students who would benefit the most from STEM participation.

Following the estimates from the discrete choice model, I present causal evidence on the heterogeneous returns to enrolling in a math-intensive major for men and women. The returns to STEM enrollment for women vary significantly by the alternative major under consideration. While STEM enrollment delivers positive returns relative to the life sciences and other majors, the returns against business and health fields are negative.[3] The returns to enrollment in math-intensive fields are lower for women than for men, partly due to sizable gender gaps in potential wages in these majors, in excess of 15 percent.[4] On the other hand, I find significant heterogeneity in these returns, such that high math ability women would largely benefit from enrolling in STEM. I also estimate the returns associated with STEM graduation after enrollment, relative to either switching fields or dropping out. These estimates acquire policy relevance in the context of initiatives aimed at supporting STEM enrollees towards degree completion. I find that all women would benefit from finishing these degrees relative to dropping out from college. Nonetheless, when compared to degree-switchers, only those above the math problem solving and self-efficacy median would benefit from finishing a STEM degree. These results suggest that broad-based STEM-inducing policy efforts would not necessarily deliver improved labor market outcomes in the early career, and as high math ability women would unambiguously benefit from enrolling in STEM, targeted policies may be preferable.[5]

Lastly, the importance of math self-efficacy in predicting women's STEM participation, coupled with the malleability of non-cognitive skills through adolescence, indicates that policies focused on boosting self-efficacy could have a sizable impact

[3]The returns to major are estimated using evidence on early-career labor market outcomes. These outcomes do not capture the full extent of lifecycle returns to college majors. These results extend Jiang (2018)'s estimated returns to STEM for women, which are estimated against non-STEM majors. I find that the pairwise comparison across different majors is key for understanding which students who are currently enrolled in other majors would benefit from starting in STEM.

[4]While these differences could be potentially explained by post-graduation occupational choices, Goldin (2014) has found sizable gender gaps within narrowly-defined science occupations.

[5]On the other hand, broad-based policies aimed at increasing women's STEM graduation rates may have significant non-pecuniary benefits. Anaya et al. (2017) have shown that girls with mothers in STEM are more likely to be employed in a math-intensive field.

on women's STEM participation rates.[6] Using the estimated model parameters, I examine the impact of a policy increasing women's self-efficacy by 0.25 standard deviations. This intervention would increase women's STEM enrollment rates by almost 20 percent relative to baseline participation rates, with larger impacts for women at the top of the problem solving distribution. This policy would also succeed in boosting graduation rates from math-intensive majors by close to 20 percent, as well. While boosting STEM participation rates may be worthwhile for non-pecuniary reasons, policymakers may also be interested in the labor market benefits arising from policy interventions. I analyze the effect of the self-efficacy intervention on women's hourly wages and find larger effects for high math ability women, for whom the returns to STEM are larger. Well-targeted skill development policies may thus help in reducing gender gaps in STEM and in narrowing gaps in early-career labor market outcomes.

This chapter contributes to an extensive literature exploring how students sort into college majors, in particular to the analysis of sorting patterns by pre-college ability. Arcidiacono (2004) has found that math ability plays a larger role in the major choice decision than verbal ability, Kinsler and Pavan (2015) find sorting into science majors based on latent math ability and Humphries et al. (2017) identify heterogeneous sorting patterns across different majors in Sweden based on grit, cognitive and interpersonal skills. These papers do not examine gender-specific sorting patterns, though a parallel set of papers has examined the factors driving

[6]An extensive literature has found non-cognitive skills to be malleable through adolescence (see summaries in Kautz et al. (2014) and in Saltiel et al. (2017).

gender differences in college majors, with Wiswall and Zafar (2014), Zafar (2013) and Wiswall and Zafar (2017) finding that differences in preferences for majors and enjoyment from studying certain fields drive the gender gap in major choices. Other papers have examined the importance of college preparation to gender gaps, with Turner and Bowen (1999) and Dickson (2010) finding that SAT scores play a small role in major gaps, and Speer (2017) finding a sizable role for ASVAB scores in understanding gaps in STEM fields. I contribute to this literature by first distinguishing between observed test scores and latent ability, as the former are influenced by background characteristics and may capture ability with measurement error. Furthermore, I identify a component of non-cognitive ability, mathematical self-efficacy, which is critical for understanding sorting patterns into STEM majors and gender gaps in this field.[7] By allowing the three components of latent ability to have gender-specific correlations, I extend the literature on unobserved heterogeneity (Heckman et al., 2016; Prada and Urzúa, 2017).

A parallel strand of the literature has analyzed the factors driving students' progression through college majors. Arcidiacono (2004), Arcidiacono et al. (2012) and Beffy et al. (2012) have estimated structural models, finding that both expected earnings and skills play a significant factor in determining initial and final major choices. Stinebrickner and Stinebrickner (2013) have found that students

[7]Previous psychology papers (Perez-Felkner et al., 2012, 2017; Nix et al., 2015) have found that self-efficacy positively affects science course completion and enrollment in science majors for both men and women. However, this work does not differentially examine how self-efficacy affects STEM enrollment and subsequent completion, nor does it distinguish between observed and latent measures. Furthermore, it does not estimate the impact of self-efficacy on gender gaps in STEM. The economics literature, on the other hand, has largely focused on analyzing the importance of test scores in gender gaps in STEM and has not considered the role of non-cognitive components.

drop out from science fields in part due to over-optimism with regards to finishing these degrees. Although the discrete choice model introduced in this chapter does not directly consider students' learning process while in college, I capture the multi-dimensional nature of ability and estimate a variety of treatment effects for quantifying the benefits from STEM enrollment and graduation. Furthermore, I estimate the model separately by gender, allowing me to understand the factors driving gender gaps in majors. A set of recent papers have presented reduced form evidence on STEM gender gaps across enrollment and graduation. Astorne-Figari and Speer (2017) and Astorne-Figari and Speer (2018) find that female STEM enrollees are more likely to switch out of these fields compared to males, and that women move to less competitive majors. Meanwhile, Kugler et al. (2017) find that women in male-dominated STEM fields are more likely to drop out in response to low grades than men. I extend this literature by exploring how different components of the latent ability vector contribute to expanding gaps at graduation and by estimating the returns arising from STEM completion after enrollment.

Lastly, I contribute to an sizable literature on the returns to majors. Altonji (1993), Rumberger and Thomas (1993), Eide (1994) and Chevalier (2011), among others, estimate linear wage models including test scores as control variables and find positive returns to STEM degrees for women, yet these papers do not correct for endogenous selection into majors. Jiang (2018) addresses this issue by presenting a discrete choice model and finds large returns to STEM degrees for women compared to non-STEM fields. Arcidiacono (2004) similarly estimates a dynamic discrete choice model in which he also finds positive returns to STEM. I build on this literature by correcting for endogenous sorting into initial and final majors, highlighting the difference in the estimated returns to majors at enrollment and graduation, analyzing how the returns to STEM vary relative to the alternative major their are compared against, and by estimating gender-specific returns.

## 4.2 Data Sources and Summary Statistics

**Data Sources**

This chapter uses longitudinal data from the Educational Longitudinal Survey (ELS) of 2002. The ELS is a nationally-representative survey of 16,700 $10^{th}$ grade students in 2002 who were interviewed, along with their parents and teachers, in the initial year, and in 2004, 2006, and 2012, when respondents had turned 26 years old. The first two surveys include detailed information on students' individual characteristics, including their race and gender, and family characteristics, including family composition, parents' educational attainment, labor market outcomes, total family income and region of residence. Critical to the analysis of sorting into majors, ELS data includes multiple test score measures. First, respondents were given a mathematics and reading test developed by the Department of Education in $10^{th}$ grade, along with a follow-up math exam in $12^{th}$ grade. ELS data also includes ACT and SAT scores for students who took these exams. Moreover, the availability of high school transcripts allows me to construct different measures of high school GPA in math and English courses.

ELS data also includes two measures of students' mathematical self-efficacy in the first and second survey. Self-efficacy is defined as an "individual's judgment about being able to perform a particular activity" (Murphy and Alexander, 2000), and Perez-Felkner et al. (2017) find that math self-efficacy positively predicts enrollment in STEM fields for both genders. The two measures are constructed directly from five questions measured on a four-point Likert scale using principal component analysis. The questions ask students to rate themselves on whether they think they can do an excellent job on math tests, understand difficult math texts, understand difficult math classes, do an excellent job on math assignments and whether the student can master math class skills.

To analyze the extent of gender differences in college majors, I restrict my sample to students enrolled in four-year college by age 20. As I impose few subsequent restrictions, the final sample includes students who do not graduate with a four-year degree, bachelor's recipients and students who have enrolled in or completed a graduate degree. I explore students' progression through college majors by using their reported major in 2006, including those who had not yet declared one, and their final major at graduation reported in the last survey. College majors are defined using a two-digit major code from the Department of Education's Classification of Instructional Programs (CIP), yielding fifty different major categories. Since working with a large number of majors is inconvenient for empirical analysis, the existing literature has often analyzed majors by aggregating them into broader categories.[8] Since Kahn and Ginther (2017) have shown that the STEM gender gap

[8]Kinsler and Pavan (2015) group majors into business, science and others, Ransom (2016) aggregates them into STEM, business, social sciences, education and others, and Jiang (2018) follows a binary STEM classification.

is largely driven by differences in math-intensive fields, I aggregate majors into five categories, which include math-intensive STEM, life sciences, Business, Health, and the remaining majors.[9]

Lastly, I analyze students' labor market outcomes using information from the third follow-up survey carried out in 2012. This survey includes detailed information on respondents' labor force participation and hourly wages. The final sample includes 4,520 respondents, with 2,010 men and 2,510 women with information for all test scores, self-efficacy measures, individual and family characteristics. Although the original ELS sample is evenly balanced between men and women, 55 percent of respondents in my sample are women.[10] In Table C.1.1, I show how the different restrictions result in the final sample used in the chapter.

**Summary Statistics**

In Table 4.1, I present descriptive statistics on the sample used in this chapter. The majority of students come from two parent families and the average surveyed parent has completed 16 years of schooling. However, male college enrollees are more likely to come from two-parent, higher income-, and higher-educated households.[11]

[9]Math-intensive STEM fields include degrees in engineering, engineering-related fields, computer science, mathematics, economics, statistics and physics. Life science degrees include majors in agriculture (and related sciences), natural resources and conservation, family science, biology and related fields and other science technologies. Business degrees includes degrees in business, management and marketing. The "Other" group includes the the following college majors: Architecture, Anthropology, Art, Art History, Communications, Criminal Justice, Education, English, History, International Relations, Journalism, Literature, Pre-Law, Political Science, Psychology, Social Work, and Sociology, among others. The Health group is largely composed of majors in Nursing, Pre-Med, Pre-Vet, Pharmacy, Health and Physical Therapy.

[10]This difference is partly explained by non-response rates in the first two survey, as restricting my sample to those who answer questions in the first survey results in a sample which is 53 percent female. Throughout the analysis, I apply sample weights to account for differential attrition by gender. The remaining difference can be explained by higher rates of college enrollment for women, as shown by Goldin et al. (2006).

[11]This difference is consistent with previous findings in Fortin et al. (2015).

In the last row, I show log hourly wages for employed men and women who are college graduates, have not gone on to graduate school and are employed at the time of the survey.[12] 93.2% and 94.3% of men and women are employed in the final survey, respectively. In this sample, the gender wage gap equals 9.5 percent, in line with the 10.4 percent wage gap for 25-29 year old college graduates in American Community Survey data.

Table 4.1 also presents evidence on students' pre-college test scores and self-efficacy measures. I find significant differences in the various math test scores available in the ELS, with men outperforming women by 0.27 standard deviations in the $10^{th}$ grade math exam developed, by 0.32 standard deviations in the $12^{th}$ grade math exam and by 0.29 standard deviations in the ACT/SAT college entrance exam. On the other hand, I examine grades in high school math courses following Niederle and Vesterlund (2010)'s insight that test scores may be partly explained by gender differences in responses to competitive pressure. I find that women earn higher grades than men by 0.14 standard deviations, suggesting that test scores overstate gaps in math ability.[13] As in Cheng et al. (2017), I find significant differences in math self-efficacy, where men's self-efficacy exceeds that of women by 0.34 standard deviations in the baseline survey and by 0.30 standard deviations in the $12^{th}$

[12]I impose this restriction as students who report having completed graduate school by age 26 may not have yet transitioned into full-time employment. As a result, their wage observations may not correctly reflect their earnings potential at the time of the last survey.

[13]These differences correspond to 0.11 and 0.12 points out of a four-point GPA scale, respectively. While I cannot directly adjust for the quality of courses taken by these students, I find no evidence of differential course-taking by gender. For instance, out of all individuals in my sample who take AP Calculus as seniors, 50.3% are women. Meanwhile, out of the sample of $11^{th}$ graders who enroll in Honors Pre-Calculus, 54.2% are women. Rothstein (2004) and Riegle-Crumb et al. (2012) find similar differences in math course grades.

grade survey. I complement this analysis by examining the relationship between math test scores and self-efficacy across the distribution in Figure 4.1. The first panel shows a strong positive relationship between the $10^{th}$ grade exam score and reported self-efficacy for both men and women, though the correlation between the two is larger for men. Furthermore, there are significant gender differences in math self-efficacy across the test score distribution, For instance, even among students in the top math test score quintile, men's observed self-efficacy exceeds that of women by 0.27 standard deviations. The second panel shows distributional differences in these two measures, where I find that the ratio of men to women in the top test score quintile is almost two-to-one, as found by Pope and Sydnor (2010), Ellison and Swanson (2010), and Guiso et al. (2008).

**Major Choices**

Table 4.2 shows the share of men and women who enroll and graduate from the five major categories defined above. There are significant differences in college major choices immediately upon enrollment. For instance, just 4.5 percent of women initially enroll in a math-intensive STEM major, relative to 17.9 percent of men. On the other hand, there are no differences in the life sciences, as seven percent of men and women start in these majors. The gender gap in STEM participation expands at graduation. Among four-year degree recipients just 4.1% of women attain a STEM degree, compared to 20% of men. This difference emerges from gaps in completion rates for both STEM enrollees and non-STEM enrollees. I explore these patterns in Figure 4.2, where I find that while 61% of male STEM enrollees subsequently complete a degree by age 26, just 44 percent of women do so. I find similar patterns among non-STEM enrollees, where 5.8 percent of men end up completing a math-intensive degree, compared to just 1.3% of women. The difference among non-enrollees plays an important role to expanding STEM gaps at graduation, as the vast majority of women and men had *not* initially enrolled in STEM.[14]

In Table 4.3, I explore sorting patterns into initial and final majors by gender, focusing on math-intensive STEM fields. I estimate a linear probability model and find math test scores and self-efficacy positively predict STEM enrollment for both women and men.[15] A one SD increase in either math component would lead to a one-third increase in enrollment rates relative to baseline participation for both genders. In the last two columns, I explore the factors driving STEM completion among students initially enrolled in these majors. For women, self-efficacy play a critical role in predicting completion, as a one SD increase in this component increases completion rates by 11.7 percentage points. This is not the case for men, for whom math test scores play a far larger role in leading to degree completion. In Table C.1.2, I expand upon this analysis by showing how math test scores and self-efficacy differ across STEM enrollees, graduates and non-completers. This analysis indicates that not only is there initial sorting into STEM, but that both men and women further sort into graduating with a STEM degree on both dimensions of math performance.[16] [17] Nevertheless, test scores cannot be considered true measures of

[14]While the literature has largely focused on differential dropout rates among STEM, I also explore the sources behind differential sorting-into-STEM rates for those who do not start in these fields.

[15]Although the estimated point estimates are larger for men, since that the baseline enrollment share for men is four times that of women the relative magnitude of the effect is similar.

[16]Astorne-Figari and Speer (2017), Astorne-Figari and Speer (2017) and Kugler et al. (2017) find that within-college factors affect switching behavior, including reaction to grades, faculty and peer composition in majors.

[17]In Table C.1.3, I examine the contribution of baseline test scores to gender gaps in STEM

ability, as they are measured with error and affected by background characteristics (Heckman et al. 2006). I present an empirical strategy which addresses this concern in Section 4.3.

Lastly, in Table C.1.4, I examine the hourly wages associated with different college majors by gender. The last three columns indicate that male STEM graduates earn the highest wages among college graduates. For women, STEM is among the best-paid fields, though health majors earn higher wages. These patterns differ when analyzing wages by initial major choice, which includes degree switchers and college dropouts. Among STEM enrollees, women earn similar wages than business enrollees and the difference against those in the life sciences falls from 25% among degree completers to 15% among the enrollee sample. As a result, an open question remains as to the magnitude of the benefits arising from STEM enrollment for women.[18] In the next section, I present a discrete choice model which accounts for endogenous sorting into initial and final college majors for both men and women. This model allows me to estimate the wage returns associated with each major separately by gender, examine how these differ at enrollment and graduation and to identify heterogeneous returns by pre-college ability.

participation. A reduced-form decomposition indicates these factors explain 15% of gaps at enrollment.

[18]While these raw differences do not represent the wage returns associated with these majors, reduced form strategies rely on selection-on-observables assumptions, which may not hold if students select majors based on their latent ability. I explore this question in Section 4.6.

## 4.3 Discrete Choice Model

This chapter estimates a sequential model of major choices and educational attainment for students initially enrolled in four-year college. This model builds on discrete choice models presented by Heckman et al. (2016), Heckman et al. (2018), Humphries et al. (2017) and Rodrıguez et al. (2017), and it follows a generalized Roy (1951) framework. It combines components used in reduced form analysis and in structural models to correct for endogenous educational choices and associated labor market outcomes. In this framework, students' decisions depend both on observed characteristics and unobserved ability.

Figure 4.3 presents the sequential decision process.[19] Upon entering college, individuals select major option $D_{m_1}$ among the set of possible majors $m_1 \in \mathcal{M}_1$.[20] $\mathcal{M}_1$ includes the five major categories defined above as well as an option to not declare a major within the first two years of enrollment. Agents then decide to continue in four-year college or drop out. Continuers are defined by $D_{d_1} = 1$. Among this group, agents select a major $D_{m_2}$ at the time of college completion from the set of five major categories, $m_2 \in \mathcal{M}_2$. Students may thus graduate from their initial major ($D_{m_1} \equiv D_{m_2}$), switch majors or drop out. Finally, agents who have completed a four-year degree have the option of completing a graduate degree $g$, $D_g = 1$.[21] Educational attainment by age 26/27 is given by $[D_{m_1}, D_{d_1}, D_{m_2}, D_g]$.

[19] Although timing is not directly considered, I am still able to capture the sequential nature of both initial and final major choices as well as college dropout.

[20] $D_{m_1}$ is a random variable which equals one if the student enrolls in major $m_1$.

[21] Given the sample size in the ELS, I restrict the graduate school decision to a binary completion decision.

The set of possible final educational attainment states is $\mathcal{S}$, agents reach one of $s \in \mathcal{S}$ and their choice is given by $D_s$. I estimate the model separately for each gender to allow for differential sorting patterns by gender and to capture gender-specific labor market outcomes. Throughout this section, the supra-index $G$ refers to an person's gender, which can either be male $m$ or female $f$.

**Initial Major Choice**

After graduating from high school and enrolling in four-year college, individuals choose an initial major based on observed characteristics and unobserved ability. The model assumes that every agent is endowed with a finite multi-dimensional vector of unobserved ability ($\boldsymbol{\theta}$), which includes cognitive and non-cognitive components of skills, known to the agent, and constant from high school through labor market entry. Since there are no direct measures of ability available, $\boldsymbol{\theta}$ is assumed to be unobserved to the econometrician, and its distribution is identified through a measurement system of pre-college test scores. By identifying the distribution of multidimensional latent ability, I can analyze how cognitive and non-cognitive components of ability affect major choices, offering an important advantage relative to structural models like Arcidiacono (2004) in which ability is unidimensional.

Let $V^G_{i,m_1}$ be the utility for student $i$ of choosing option major $m_1$ from the set of all possible initial choices, $\mathcal{M}_1$. $V^G_{i,m_1}$ represents an approximation of the value of each major for individual $i$ and it incorporates agents' perceived economic returns to each major and non-pecuniary tastes, but does not impose any direct structure on the decision-making process by postulating preferences and/or information sets. As a result, students are allowed to make irrational decisions, or even mistakes, which may be subsequently changed in their final choice, as additional information is revealed.[22] $V_{i,m_1}^G$ depends on both observed characteristics and unobserved ability. It is given by:

$$V_{i,m_1}^G = \beta_{m_1}^G X_{i,m_1}^G + \alpha_{m_1}^G \boldsymbol{\theta_i}^G + \varepsilon_{i,m_1}^G \quad for\ m_1 \in \mathcal{M}_1 \tag{4.1}$$

where $X_{i,m_1}^G$ represents the vector of exogenous characteristics, $\boldsymbol{\theta_i}^G$ captures the vector of latent ability, and $\varepsilon_{i,m_1}^G$ is the error term. Conditional on individual observed and unobserved characteristics, major choices are unordered. More precisely, individuals choose the major that yields the highest utility such that $D_{i,m_1}^G = 1$ if $m_1 = \text{argmax}_{m_1 \in \mathcal{M}_1} \left\{ V_{i,m_1}^G \right\}$.

**Final Major Choice**

After initially enrolling in major $D_{i,m_1}^G$, students either continue in college through graduation or drop out. Continuers select their field at graduation $m_2$ from the set of possible options $\mathcal{M}_2$. As noted above, the second major choice may involve continuing with the same major or switching fields.[23] This step encompasses two decisions. First, $V_{i,d_1,m_1}^G$ represents an approximation to the net utility associated with continuing in four-year college after enrollment. $V_{i,d_1,m_1}^G$ depends on observed

[22] Allowing agents to not declare a major upon at college entry fits in with this consideration, as they may prefer to wait to declaring until acquiring additional information on different fields.

[23] An extensive structural literature has analyzed the factors behind changing major choices within college (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2013; Wiswall and Zafar, 2014), including learning about academic ability, and changing expectations, among other reasons. As this model is not structural, I do not impose any specific structure on the learning process associated with final major choices. Individuals' choices at each decision node may be influenced by the realization of the error term, yet these shocks do not persist through future decisions. Unlike this literature, the discrete choice model presented in this chapter instead focuses on understanding the importance of cognitive and non-cognitive ability in college major choices. I discuss potential biases arising from ability updating in Appendix D.1.

and unobserved characteristics and is defined as follows:

$$V^G_{i,d_1,m_1} = \beta^G_{d_1,m_1} X^G_{i,d_1,m_1} + \alpha^G_{d_1,m_1} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,d_1,m_1} \tag{4.2}$$

Students continue in college if $V^G_{i,d_1,m_1} > 0$. $D^G_{i,d_1} = 1$ equals one for continuers. Let $V^G_{i,m_2,d_1,m_1}$ be the utility for individual $i$ of choosing option major $m_2$ from the set of all possible final choices, $\mathcal{M}_2$, given their initial choice $m_1$. I allow the utility from the major at graduation to depend on the initial choice $m_1$ as individuals may derive further pecuniary and non-pecuniary benefits from completing the major they had initially enrolled in. $V^G_{i,m_2,d_1,m_1}$ depends on a set of observed characteristics and the vector of unobserved ability. It is specified as follows:

$$V^G_{i,m_2} = \beta^G_{m_2} X^G_{i,m_2} + \alpha^G_{m_2} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,m_2} \quad for\ m_2 \in \mathcal{M}_2 \tag{4.3}$$

$X^G_{i,m_2}$ is a vector of exogenous characteristics, $\boldsymbol{\theta_i}^G$ represents latent ability endowments, and $\varepsilon^G_{i,m_2}$ is the error term. $D^G_{i,m_2}$ is a dummy variable representing the final major choice, given by the major $m_2$ yielding the highest utility.

**Graduate School**

Since a sizable share of ELS graduates attain a graduate degree by age 26, I include this decision margin as part of the analysis of educational attainment and I examine how men and women sort into graduate education based on their ability endowments. I model this decision as a binary probit model, where students decide whether to complete a graduate degree by age 26 or not.[24] $V^G_{i,g,m_2,d_1,m_1}$ represents an

[24]As ELS data does not provide information on subsequent educational attainment, I classify

approximation to the utility associated with graduate school choices and it depends on the history of major choices to take into account that, for instance, graduate school may be more appealing for a STEM graduate relative to a life sciences graduate. The structure behind this decision is similar to that of major choices, depending on observed characteristics and unobserved ability (for notational simplicity, I omit dependence on prior choices):

$$V_{i,g}^{G} = \beta_{g}^{G} X_{i,g}^{G} + \alpha_{g}^{G} \boldsymbol{\theta_i}^{G} + \varepsilon_{i,g}^{G} \tag{4.4}$$

As in the previous decision, $X_{i,g}^{G}$ is a vector of exogenous characteristics, $\boldsymbol{\theta_i}^{G}$ represents latent ability endowments, and $\varepsilon_{i,g}^{G}$ is the error term. $D_{i,g}$ is a dummy variable which equals one if the person chose to finish a graduate degree by the last survey round. All in all, the combination of student $i$'s educational decisions $[D_{i,m_1}, D_{i,d_1}, D_{i,m_2}, D_{i,g}]$ implies that students reach one of $s \in \mathcal{S}$ final educational attainment states, captured by the dummy variable $D_{i,s}$.

**Labor Market Outcomes**

The labor market outcome of interest in this chapter is a person's hourly wage at age 26, given by $W_{i,s}$, corresponding to student $i$'s educational attainment. Potential wages are also also determined by observed characteristics and unobserved abilities and are defined as:

$$W_{i,s}^{G} = \beta_{s}^{G} X_{i,s}^{G} + \alpha_{s}^{G} \boldsymbol{\theta_i}^{G} + \varepsilon_{i,s}^{G} \;\; \forall \; s \; \in \; \mathcal{S} \tag{4.5}$$

students by their highest degree attained in the last survey round.

where $X_{i,s}^G$ represents a vector of exogenous control variables determining hourly wages and $\varepsilon_{i,s}^G$ is the associated error term, which is assumed to be uncorrelated with observed and unobserved characteristics. Furthermore, hourly wages are only observed for individuals who choose to participate in the labor market, $D_{s,e}$.[25] This decision follows the structure of all previous decisions and the latent utility associated with working is:

$$V_{i,se}^G = \beta_{se}^G X_{i,se}^G + \alpha_{se}^G \boldsymbol{\theta_i}^G + \varepsilon_{i,se}^G \ \ \forall \ s \ \in \ \mathcal{S} \tag{4.6}$$

where $X_{i,se}^G$ represents a vector of exogenous control variables determining hourly wages and $\varepsilon_{i,se}^G$ is the associated error term. Equations (4.1)-(4.6) imply that unobserved abilities $\boldsymbol{\theta}$ affect labor market productivity, initial and final major choices, and graduate school decisions. Equation (4.5) describes individual $i'$'s wages in each final attainment node, a critical component for understanding the returns to college majors.

**Measurement System**

Since latent ability $\boldsymbol{\theta}$ drives the endogeneity of decisions and generates all cross-correlations of outcomes and choices conditional on $X_{i,m_1}$, $X_{i,m_2}$, $X_{i,g}$, $X_{i,s}$, and $X_{i,se}$, identifying its distribution is of paramount importance in this model.[26] Since $\boldsymbol{\theta}$ is unobserved to the researcher, I follow an extensive literature and allow for

[25] Hourly wages are only modeled for individuals who have not completed graduate school ($D_g = 0$). $D_{se} = I\left[V_{se} > 0\right]$ is a dummy variable which equals one for individuals who work in the final survey round.

[26] The rest of the unobserved components of the model are independent across educational choices and and labor market outcomes. Formally, this means that $\varepsilon_{i,m_1} \perp \varepsilon_{i,m_j} \forall m_1, m_j \in \mathcal{M}_1$, $\varepsilon_{i,m_2} \perp \varepsilon_{i,m_k} \forall m_2, m_k \in \mathcal{M}_2$, $\varepsilon_{i,s} \perp \varepsilon_{i,n} \forall m, n \in \mathcal{S}$, $\varepsilon_{i,se} \perp \varepsilon_{i,je} \forall s, j \in \mathcal{S}$, and $\varepsilon_{i,m_1} \perp \varepsilon_{i,m_2} \perp \varepsilon_{i,s} \perp \varepsilon_{i,ke} \forall m_1 \in \mathcal{M}_1, \forall m_2 \in \mathcal{M}_2, s \in \mathcal{S}, \forall k \in \mathcal{S}$.

$\boldsymbol{\theta}$ to be proxied by multiple measures of pre-college test scores in mathematics and reading as well as by measures of math self-efficacy (Carneiro et al., 2003; Hansen et al., 2004; Heckman et al., 2006; Urzua, 2008). Identifying the distribution of $\theta$ through a measurement system also allows me to correct for measurement error in observed test scores, as each particular test score measures latent ability with error. In fact, separate estimation by gender allows me to account for the existence of differential measurement error in test scores for men and women (Cattan, 2013). I posit a linear model in which test scores are modeled as a linear outcome determined by latent ability $\boldsymbol{\theta}$, individual and family characteristics.[27]

I observe nine different test score measures. To determine the number of factors and the structure of the measurement system, I perform an exploratory factor analysis (EFA) using the nine observed measures. Assuming orthogonal factors, exploratory factor analyis yields four factors with positive eigenvalues for women (3.81, 0.95, 0.68, 0.02) and similar values for men (3.95,1.04,0.64,0.03). Motivated by Cattell (1966)'s scree test, these results indicate that at three factors are needed to explain the relationship between the observed measures. In Figure F.2.1, I show the estimated coefficients associated with each factor by gender. These results indicate that all nine measures load positively on the first factor, with significantly larger coefficients on the math test score measures as well as on math grades. Meanwhile, the second factor loads strongly on the two math self-efficacy measures and on math GPA and the third factor is only relevant in the English/Reading test scores and in

[27] As an an extensive literature has shown the importance of family, cultural and social factors in determining the evolution of ability through childhood, I interpret the components of $\theta$ to be fixed by the time of college enrollment, but not fixed from birth or indicative of gender differences in inherited ability.

high school English grades. These results suggest that the first factor largely reflects students' math ability, the second factor captures their math self-efficacy and the third factor identifies their reading ability.

Following the insights from exploratory factor analysis, I posit the existence of three components of latent ability, which I define as math problem solving ability $\theta_C$, math self-efficacy $\theta_{SE}$, and reading ability $\theta_R$.[28] Given the estimated loadings shown in Figure F.2.1, I allow for all math and reading test scores, as well as math self-efficacy and high school grades to be a function of math problem solving ability. Since Borghans et al. (2008) find GPA to be a function of both cognitive and non-cognitive skills, I allow math GPA measures to also be a function of math self-efficacy, for self-efficacy measures to also depend on $\theta_{SE}$. Meanwhile, reading test scores and English high school grades depend on on the latent reading ability component.

The model for math test scores ($C^G_{i,j}$) can be expressed linearly as follows:

$$C^G_{i,j} = \beta^G_{C_j} X^G_{i,C_j} + \alpha^G_{C_j} \theta^G_{C,i} + \varepsilon^G_{i,C_j} \tag{4.7}$$

[28]Problem solving ability is defined as the "process in which an individual uses previously knowledge ... to satisfy the demands of an unfamiliar situation" (Krulik and Rudnick, 1989). This component of ability has been previously considered in the context of math problem solving (Grattoni, 2007). Carneiro et al. (2003) show the identification of the distribution of unobserved ability requires the at least seven test scores in a model with three components. This requirement is met in this sample.

Similarly, the model for math GPA ($G^G_{i,1}$) and math self-efficacy ($SE^G_{i,n}$) is given by:

$$G^G_{i,1} = \beta^G_{G_1} X^G_{i,G_1} + \gamma^G_{G_1}\theta^G_{SE,i} + \alpha^G_{G_1}\theta^G_{C,i} + \varepsilon^G_{i,G_1} \tag{4.8}$$

$$SE^G_{i,n} = \beta^G_{SE_n} X^G_{i,SE_n} + \gamma^G_{SE_n}\theta^G_{SE,i} + \alpha^G_{SE_n}\theta^G_{C,i} + \varepsilon^G_{i,SE_n} \tag{4.9}$$

Finally, the model for English GPA and for reading test scores ($R_k$) follows:

$$R^G_{i,k} = \beta^G_{R_k} X^G_{i,R_k} + \eta^G_{R_k}\theta^G_{R,i} + \alpha^G_{R_k}\theta^G_{C,i} + \varepsilon^G_{i,R_k} \tag{4.10}$$

Across equations (4.7)-(4.10), $\boldsymbol{X}$ represents a vector of exogenous control variables and $\boldsymbol{\varepsilon}$ represents the error term. In Appendix D.1, I show how the measurement system secures the identification of the distribution of latent ability.[29] All error terms, $\varepsilon_{i,C_j}$, $\varepsilon_{i,G_1}$, $\varepsilon_{i,SE_n}$, and $\varepsilon_{i,R_k}$ are mutually independent, independent of $\theta_C$, $\theta_{SE}$, $\theta_R$ and independent of $\boldsymbol{X}$.

Early papers in this literature assumed the components of latent ability to be independent from each other (Hansen et al., 2004; Heckman et al., 2006). Nonetheless, given the high correlation present between observed math test scores and self-efficacy, the two components of latent math ability may be correlated as well. As a result, I follow two recent papers (Heckman et al., 2006; Prada and Urzúa, 2017), and

[29] As shown in Heckman et al. (2006), since there are no intrinsic units for the latent ability measures, one coefficient devoted to each component must be normalized to unity to set the scale of each component of ability. Therefore, for some math test score measure $j$, self-efficacy measure $n$ and reading test score $k$, I set $\alpha_{C_1} = 1$, $\gamma_{SE_1} = 1$ and $\eta_{R_1} = 1$. The results are robust to different normalizations. I note that the measurement system presented above can be extended to allow for $\theta_{SE}$ to affect performance in one math test score. I separately estimated a parallel measurement system allowing for $\theta_{SE}$ to affect Math SAT performance and found that latent math self-efficacy explained less than 0.2% of the variance in math SAT performance.

allow for the latent ability components to be correlated. In Appendix D.1, I discuss the assumptions required to identify the correlation between ability components. This chapter extends the literature on latent factors by allowing the correlation between latent ability components to be gender specific.

**Identification**

The identification of the joint distribution of counterfactual outcomes and educational choices follows from formal arguments presented in Heckman and Navarro (2007) and Heckman et al. (2016). The model is identified through a combination of a matching-on-unobservables assumption and node-specific exclusion restrictions. First, I secure the identification of the distribution of unobserved ability $\boldsymbol{\theta}$ through the measurement system in equations (4.7)-(4.10), which requires the normalization of one loading in each latent component and for $\boldsymbol{\theta}$ to be orthogonal to $\boldsymbol{X}$ and $\varepsilon$. The formal argument is laid out in Appendix D.1. Second, a conditional independence assumption implies that all college major choices, graduate school decisions and labor market outcomes are independent conditional on all observed characteristics and unobserved ability components. This assumption can also be understood as a "matching" assumption, which extends reduced form approaches by allowing for matching on unobserved ability, as well as on observed characteristics.

The model can be identified solely through the conditional independence assumption, yet Heckman et al. (2016) note introducing exclusion restrictions at each decision node allows for an identification at infinity argument. Finding economically-meaningful shifters of initial and final major choices is challenging, especially in the U.S. context, where college majors are largely priced uniformly. In the first decision node, I use the share of students enrolled in student $i$'s local four-year college(s) who completed major $m_1 \in \mathcal{M}_1$ as an exogenous shifter.[30] This share is gender-specific and this variable may affect students' major choices through a role-modeling effect, as they reside in areas with a varying shares of college graduates in each specific major $m_1$. For the next two educational decisions, I follow Heckman et al. (2016) and Heckman et al. (2018) and use local unemployment rates by major as exogenous shifters, following the intuition that local major-specific unemployment rates may affect students' perceived benefits arising from different choices. For the dropout decision, I use the local unemployment rate for own-gender college graduates, given by the commuting zone of residence in the third survey round. For the final major choice decision, I use local unemployment rates by major. For the employment decision, I consider local unemployment rates by college major, as well, but defined at the students' commuting zone of residence in the final survey round.[31] Lastly, for the graduate school decision, I use the share of college graduates aged 25-34 who have also obtained a graduate degree in person $i$'s commuting zone of residence in the final survey round. Table C.1.5 shows the variables used in the implementation of the model. These variables are the same for both genders.

**Implementation**

To define the sample likelihood, I collect all exogenous controls in the educa-

[30]This variable is constructed by identifying students' commuting zone of residence in the baseline survey and matching it to IPEDS data indicating the number of students completing major $m_1$ in the colleges in the respective commuting zone. The IPEDS data used to create this variable is from 2000 to ensure that students included in the survey are not captured in the average share of students choosing a particular major. The unemployment rate for the undeclared major option is the average in each commuting zone for college graduates.

[31]I construct these variables from American Community Survey 2010 data (obtained from public-use IPUMS NHGIS data).

tional choice and outcome equations in the vector $\mathbf{X}_i$ and the vector of test scores in $\mathbf{T}_i$. Given the independence assumptions invoked above, the likelihood for a set of $I$ individuals is given by:

$$\begin{aligned}\mathcal{L} &= \prod_{i \in I} \iiint f(W_{is}, D_{i,se}, D_{i,g}, D_{i,m_2}, D_{i,d_1}, D_{i,m_1}, \mathbf{T}_i | \mathbf{X}_i, \theta_C, \theta_{SE}, \theta_R) \\ &\quad dF_{\theta,C}(.) dF_{\theta,SE}(.) dF_{\theta,R}(.)\end{aligned}$$

I assume the error terms in the measurement system, initial major choice, drop out decision, final major choice, graduate school decision, employment decision and in the wage equation are normally distributed. The initial and final major decisions are estimated with a multinomial probit. The employment, graduate school and college dropout decisions are estimated using a probit model.[32]

The model is estimated separately by gender and to estimate the density function of each unobserved factor, I use flexible distributional assumptions. I initially assume that the vector of unobserved ability is an independent random variable with mean zero. I later relax this assumption to examine gender differences in latent ability. I approximate the distribution of each ability component $k \in \{C, SE, R\}$ using a mixture of two normal distributions with means $(\mu_{1,k}, \mu_{2,k})$, probabilities $(p_{1,k}, p_{2,k})$,

[32]In some initial majors, few individuals switch into every possible option in $\mathcal{M}_2$. For instance, no women switch from the life sciences to business. As a result, I impose a restriction similar to Cameron and Heckman (2001) such that for individuals starting in STEM, life sciences, business and health, $\mathcal{M}_2$ includes remaining in the major or switching to any other major. For those starting in Other majors or non-declared, $\mathcal{M}_2$ includes the full set of majors. This assumption leads to only 30 men and 40 women being misclassified. Note that in the final educational states $s \in \mathcal{S}$ which include having completed a graduate degree by age 26, neither the employment decision nor hourly wages are considered.

with $p_{1,k} + p_{2,k} = 1$, and variances $((\sigma_{1,k})^2, (\sigma_{2,k})^2)$ as follows:

$$\theta_k \sim p_{1,k}\ N(\mu_{1,k}, (\sigma_{1,k})^2) + p_{2,k}\ N(\mu_{2,k}, (\sigma_{2,k})^2)$$

Given the numeric complexity in estimating the likelihood, the model is estimated by Markov Chain Monte Carlo (MCMC) as in Hansen et al. (2004), and Heckman et al. (2006), among others.[33] Using the estimated model, I simulate 100 samples from the original sample, such that each new sample comes from a different draw from the posterior of distribution of structural parameters, yielding a total of 451,000 observations.

Table C.1.5 shows the variables used in the implementation of the model. These variables are the same for both genders. Meanwhile, Table C.1.6 presents the estimated coefficients for the choice equations and labor market outcomes for women initially enrolled in STEM. As shown in the Table, the coefficients on the various exclusion restrictions follow the expected sign, yet are of varying statistical significance. As such, I remark the importance of the conditional independence assumption, which ensures model identification.

**Goodness of Fit**

To examine the validity of the discrete choice model in matching observed educational choices, I conduct various goodness of fit tests. In Panel A of Table

[33]Using a vector of initial parameters from the transition kernel, the Markov Chain is generated according to the Gibbs sampler, such that as $n \to \infty$, the limiting distribution is the posterior. Once convergence is achieved, I make 1,000 draws from the posterior distribution of estimated model parameters to compute the mean and the standard errors of the parameters of interest. For more details, see Hansen et al. (2004) and Heckman et al. (2006).

C.1.7, I contrast workers' observed initial major choices by gender against those simulated in the model. The model accurately predicts major choices by gender, with the majority of students in 'Other' majors and men outpacing women in math-intensive majors. I confirm this result with a $\chi^2$ test of the equality of means, finding that observed and simulated major choices are not statistically different for either gender. In Panel B, I compare the observed and simulated final major for students who started in a STEM field. Again, I find no significant differences for either gender. Finally, in Panel C, I explore the final major for students who had not initially declared a major and find that the observed and simulated transition shares are well-predicted. In Table C.1.8, I present evidence on the goodness of fit for the employment decision and log hourly wages in each higher education for women. The model predicts well the employment decision and the mean in hourly wages across choices, except in two nodes, though the differences are only significant at the 5 percent level.

## 4.4 Model Results: Latent Ability

**Measurement System**

Tables B.1.3 and B.1.4 present the estimated coefficients from equations (4.7)-(4.10) on the nine observed test score measures for both genders. Men and women from two parent families and those with more educated parents are more likely to score higher on the various test score, GPA and self-efficacy measures. This component is relevant to the analysis of gender gaps in test scores, as male college enrollees come from more educated families relative to women. For both men and women, having a parent in a STEM occupation increases test score performance, both in math and reading, yet for women, the point estimates associated with having a mother in STEM are generally larger than those whose father is in STEM. Women's math self-efficacy is positively affected by having a mother in STEM, as is the case for men, for whom the magnitudes are smaller. Moreover, the positive factor loadings across the measurement system indicate that observed measures are partly determined by latent ability. Lastly, the magnitude of the factor loadings is similar across genders.[34]

To understand the relative contribution of students' background characteristics and their latent ability vector for each test score, I present a variance decomposition of the measurement system in Figure 4.4. For the math and reading test scores, the share explained by observable characteristics reaches 10 percent for both men and women. Observed characteristics explain a much smaller share of the variance in math GPA and self-efficacy, indicating that despite the positve loading on 'Mother in STEM,' students backrground characteristics do not explain a sizable share of their self-efficacy in mathematics. On the other hand, this exercise confirms the critical role of latent ability for explaining the variance in the observed measures. For the three math test scores, the latent math problem solving component explains between 60 and 75 percent of the variance in test scores. Meanwhile, a large share of the variance in observed self-efficacy measures is explained by the latent self-efficacy

[34]In the context of cross-gender comparisons of factor distributions, configural invariance requires for observed measures to be dedicated to the same unobserved ability component for both men and women. Following Gregorich (2006), Cattan (2013) argues that similar point estimates in the loadings structure across genders can be interpreted as evidence of configural invariance.

component, which accounts for 35-70% of the variance, whereas the problem solving factor explains less than 10% of the variance in these measures. Moreover, one-third of the variance of high school math GPA is explained by the problem solving component with less than 10% explained by self-efficacy, confirming Borghans et al. (2008) finding that GPA is a function of both cognitive and non-cognitive skills. Lastly, around 25-50% of the variance in reading/English test scores is explained by the reading factor, with an additional 5-15% being explained by the math problem solving factor. This evidence supports the argument that test scores cannot be equated with latent ability, as they are direct functions of background characteristics and capture distinct components of the ability vector. As a result, any empirical strategy which equates math test scores with latent math ability should be interpreted with caution.

**Gender Differences in Latent Ability**

In Section 4.3, I had initially assumed that the mean of each factor for both males and females equalled zero. To identify gender differences in the means of unobserved abilities, I extend Urzua (2008)'s method to accommodate a measurement system in which observed measures depend on multiple measures of ability. Given the variance decomposition presented in Figure 4.4, I assume that gender differences in average math test scores only contribute to gaps in the mean of latent math problem-solving ability and that differences in math GPA reflect gender gaps in the mean of the problem solving factor.[35] In Appendix D.2, I present the formal

[35] As discussed in Appendix D.2, the results are robust to different assumptions. I assume that gender gaps in the average reading test score and in the English SAT score explain average differences in the reading factor. This assumption is restrictive, as Figure 4.4 shows these two measures loading on the problem solving factor as well. I do not focus much attention in Sections 4.5 and

argument behind the identification of gender differences in the latent ability means after imposing these assumptions.

Estimating the model separately by gender allows me to recover gender-specific distributions of unobserved ability and correlations between each ability component. In the first panel of Table 4.4, I present summary statistics on each component of latent ability by gender. The gap in the math problem solving component equals 0.16 standard deviations, which is significantly smaller than the average gap of 0.29 SDs in math test scores. This difference is explained both by the fact that college-enrolled men come from more educated families and by high school GPA loading positively on the problem solving component. I remark that empirical analyses which equate gaps in math test scores with gender differences in math skills vastly overstate the gap.

I also find significant average differences in the latent math self-efficacy component, in the range of 0.151 standard deviations. Finally, the gender gap is reversed in the reading component, with women surpassing men by 0.13 standard deviations, on average. Figure C.1.1 shows the marginal densities of each component of math ability by gender. The distribution of women's problem solving ability is dominated by the male distribution, confirming average gaps presented in Table 4.4. These differences emerge across the latent ability distribution, with men making up 62% of the top problem solving ability decile, far exceeding their 44.5% share in the sample. Panel B shows the marginal distribution of the self-efficacy component, where again

4.6 exploring gaps in reading ability, however. These assumptions do not affect the interpretation nor the magnitude of subsequent empirical analysis.

the distribution of women's self-efficacy is dominated by that of men.

### Correlation of Latent Abilities

I find a large and positive correlation across the three ability components for men and women. Most important to the analysis of sorting into STEM, however, is the correlation between the two math ability components, which equals 0.56 for men, far surpassing the 0.47 correlation for women. In Figure 4.5, I present the joint distribution of the two math ability factors by gender, which shows the high correlation between these components. For instance, a large share of men and women in their own-gender's top decile of the math problem solving component are also in their the top self-efficacy decile. Nonetheless, interesting gender differences emerge: while 32% of men in the top problem solving decile are also in the top self-efficacy decile, the equivalent share is 27% for women. Furthermore, just 15% of men in the top math decile are below the median of the self-efficacy component, yet this is the case for 23% of women, confirming an over-representation of high-skilled women who aren't confident in their math ability. In this context, Carlana (2018) finds that teachers' gender stereotypes lower girls' subsequent performance and self-confidence in math, indicating that the lower correlation in latent math ability for girls may be a function of external influences.[36] In the next section, I analyze the nature of

[36]In ongoing work, I explore the timing at which these differences may emerge. Using ECLS-K data, I find significant gender gaps in children's self-reported math competence/interest in the third and fifth grade, in the range of 0.15-0.20 standard deviations. At the same time, the correlation between math test scores and math interest in both grades is lower than 0.20, with small differences indicating a larger correlation for boys than for girls. I complement this analysis with data from the High School Longitudinal Study (HSLS), where, among ninth graders, I find that a math test score has a correlation of 0.35 with girls' math identity, but of 0.41 among boys. This preliminary analysis suggests that an important component of math skill development among children is the progressive reinforcement of math performance and perceived math interest, though the relationship between these components is weaker for girls than it is for boys.

sorting into majors by pre-college skills for men and women.

## 4.5 Model Results: College Major Choices

### 4.5.1 Enrollment Decisions

Using the estimated model parameters, I examine how students sort into initial majors based on their pre-college latent ability vector. Men and women sort into math-intensive majors based on both components of mathematical ability. The cumulative ability distribution for students enrolled in STEM majors stochastically dominates the distribution of those in other majors. For instance, women in STEM have problem solving ability and self-efficacy that is 0.39 and 0.34 standard deviations higher than that of those enrolled in business-related majors, respectively. I find the same pattern for men, with a difference of 0.48 and 0.44 standard deviations in problem solving and self-efficacy between students enrolled in STEM and business. For students in other fields, sorting patterns are less clear, though both men and women in 'Other' majors rank the lowest in both components of latent math ability.[37]

Figure 4.6 shows the relationship between both components of math ability and STEM enrollment. The left panel shows that women who are in the top joint decile of problem solving and self-efficacy are far more likely to start in STEM (13 percent) than those in the middle joint decile (3.5 percent). Self-efficacy plays a

[37]Sorting patterns based on reading ability are less stark. For instance, women in the life sciences and in STEM have the highest reading ability, yet outpace those in between Health, Other fields and non-declared students by just 0.2 SD. I find a similar pattern for men, though those in STEM have higher reading ability than students in the life sciences.

critical role in this decision: among women in the top problem solving decile, moving from the bottom self-efficacy decile to the top one increases STEM participation rates by 11 percentage points. This result gains importance in the context of the lack of women at the top of the joint math ability distribution shown in Figure 4.5. Figure 4.6 also shows that STEM participation is a non-linear decision for women. For instance, a woman in the bottom problem solving decile who moves from the bottom self-efficacy decile to the top one would only increase her expected STEM participation by 2.1 percentage points, less than one-fifth of the corresponding effect for a student in the top problem solving decile. Despite the pronounced sorting patterns on math ability, a sizable share of high-achieving women instead enroll in other majors. Among women in the top decile of the joint math distribution, 17% enroll in the life sciences, 33% in a major in the 'Other' category, and 11.6% do so in a health-field.

The panel on the right of Figure 4.6 shows sorting patterns for men on both dimensions of math ability. There are significant gender differences in the share enrolled in STEM, both in levels and in slope. For instance, 14.9% of men in the middle of the joint math distribution initially enroll in STEM, which exceeds enrollment rates for women in the top of their gender's joint distribution of math ability (13 percent). Upwards of 41 percent of men in the top joint decile begin in STEM, almost tripling women's participation in the equivalent skill ranking. In fact, for men in the top decile of their gender's joint math ability distribution, 11% in the life sciences, 2.6% in health and 19% in other fields. The largest gender differences among high-ability students appear in STEM enrollment and in health-related fields, and these differences persist across the math ability distribution.

An alternative identification strategy could have analyzed semi-parametric sorting patterns using observed math test scores as proxies for ability. I present these results in Table B.1.5 and Figure C.1.2, which show that reduced-form analysis cannot correctly capture the importance of non-linearities in math ability in determining STEM enrollment patterns for men and women, as test scores measure latent ability with significant error.

### 4.5.2 Final Major Choices

STEM graduation rates are a combination of completion rates among students who started in this field and switching-into-STEM rates among students in other majors. Among STEM enrollees, women were are less likely to complete a degree in this field than men. The left panel of Figure 4.7 shows heterogeneous completion rates for women who started in STEM. Both components of math ability affect the likelihood of degree completion among female STEM enrollees. For instance, a woman in the middle quintile of the marginal problem solving distribution has a 35 percent chance of eventually graduating with a STEM degree, yet this probability increases to 58 percent for those in the top quintile. Non-cognitive skills are similarly important: 22 percent of STEM-enrollees in the bottom self-efficacy quintile complete a degree, rising to 35 percent for those in the middle quintile, and reaching upwards of 56 percent of women in the top self-efficacy decile. The joint skill distribution presents a similar story, as 63 percent of women in the top joint quintile graduate with a

STEM degree, yet this share drops to below 28 percent for those in the top problem solving quintile and in the bottom of the self-efficacy distribution. These patterns are strikingly different for men, for whom self-efficacy plays a far smaller role in determining degree completion. While 55% of STEM enrollees in the bottom quintile complete a degree after enrollment, this share rises only slightly to 59% and 64% for those in the middle and top self-efficacy quintiles, respectively. This result shows the importance of considering how different margins of ability differentially affect men and women's progress through majors in college. In particular, non-cognitive skills play a critical role for women's exit from STEM, yet this margin has not received much attention in the existing literature. In fact, a shortfall in math self-efficacy may explain female dropout from STEM fields in response to low grades (Kugler et al., 2017).

In Section 4.2, I had shown that the small share of women switching into STEM from other majors vis-a-vis male switching rates played an important role in expanding gender gaps in STEM majors at graduation. In Figure C.1.3, I examine sorting-into-STEM patterns for women who had not initially enrolled in these fields. While the average switching-into-STEM-rate is small (one percent), there is significant sorting on the problem solving ability component: 0.8 percent of women in the middle decile end up completing a STEM degree, which is far lower than the 2.7 completion rate for those in the top decile. Sorting on the self-efficacy component is less prevalent, where, on average, 1.7 percent of women in the top of the distribution complete a STEM degree compared to 1.2 percent of those in the middle decile. This result differs from the importance of self-efficacy for female STEM enrollees in completing those degrees. A potential story behind this result is that being exposed to difficult math-intensive classes early on in college requires students to persevere by relying on their non-cognitive skills. On the other hand, as women in other majors do not face equivalent challenges, they choose to switch into STEM largely based on their math problem solving ability.

In Table B.1.6, I expand upon these results by analyzing the productivity of different components of the latent ability vector in leading to college and STEM completion among students enrolled in STEM and in other fields. I estimate the impact of a one standard deviation increase in each of the ability components. Math problem solving ability has a sizable impact on STEM completion rates for female STEM enrollees, increasing completion rates by 13.6 percentage points, with similar impacts for men. Confirming the results in Figure 4.7, the returns to self-efficacy for women are large and significant, increasing completion rates by 10 percentage points. On the other hand, the effect for men is not different from zero. For students enrolled in other fields, both components of math ability have an negligible effect on subsequent STEM completion.

### 4.5.3 Closing Gender Gaps in STEM

Since women sort positively into STEM based on both components of math ability, and as I had found gender gaps in math ability in Section 4.4, I examine whether closing gaps in math skills could lead to increased female enrollment in STEM. Following equation (4.1), women's participation in any initial major $m_1$ can be generally expressed as: $D^f_{m_1} = g(X^f_{m_1}, \boldsymbol{\theta}^f)$. This expression can be used to compute the contribution of observed and unobserved factors to women's enrollment in any major $m_1$. I analyze how closing distributional gender gaps in math could increase women's STEM enrollment in:

$$D^f_{m_1,\Delta} = g(X^f_{m_1}, \boldsymbol{\theta_C}^m, \boldsymbol{\theta_{SE}}^m) \quad \forall \;\; m_1 \in \mathcal{M}_1 \tag{4.11}$$

where $\boldsymbol{\theta_k}^m$ represents the male distribution of the $k^{th}$ component of latent ability.[38] I present the results in Table B.1.7. Compensating women with men's marginal distribution of unobserved math ability would only increase the share of women choosing math-intensive fields from 4.5 percent to 5.3 percent, with a corresponding increase to 5.3 percent for the self-efficacy compensation. Furthermore, closing the distributional gap in both dimensions would increase women's STEM enrollment to 6.2 percent, making up 14 percent of the initial enrollment gap in this field.

As enrollment does not imply completion, I also examine how eliminating distributional gaps in math skills would affect women's STEM completion rates. Following equation (4.11), this effect is given by:

$$D^f_{m_2,\Delta} = g(X^f_{m_2}, \boldsymbol{\theta_C}^m, \boldsymbol{\theta_{SE}}^m) \quad \forall \;\; m_2 \in \mathcal{M}_2$$

I present the results in the second panel of Table B.1.7. Similar to the estimated impact on initial STEM enrollment, eliminating distributional gaps in the prob-

[38]Following Kahn and Ginther (2017) 's review of the emergence of gender gaps in math test scores, this exercise can be considered an approximation of the effect of holding gender math achievement gaps constant from elementary school through high school.

lem solving dimension would increase women's estimated completion rates in math-intensive fields from 2.9 percent to 3.8 percent, whereas the equivalent increase in self-efficacy would yield a corresponding increase to 3.7 percent. The elimination of gender gaps in both dimensions of math ability would increase the share of women completing a STEM degree to 4.7 percent, thus closing almost 15 percent of the gender gap in math-intensive STEM majors.

The existing literature on this topic has found that gender differences in preferences can explain a sizable share of gaps in STEM fields. For instance, Zafar (2013) has found that beliefs about enjoying coursework explain 50 percent of gender gaps in engineering majors. The results presented in this section indicate that gender differences in math skills do not explain a majority of the gap in STEM enrollment and graduation rates, yet the role of skills is non-negligible, as distributional gaps in math ability explain almost fifteen percent of gaps in STEM enrollment and graduation rates. Furthermore, I have found that women with higher endowments in both dimensions of math ability are more likely to enroll and persist in these fields. I lastly note that the interaction between ability and preference formation remains an open question, as, for instance, preferences among college enrollees may be a function of both cognitive and non-cognitive skills while in high school.[39] The results presented in this section complement the existing literature analyzing the factors which drive women's participation in STEM. Nonetheless, the evidence on whether

[39]In ongoing work, I take advantage of HSLS data to analyze how math test scores and self-efficacy in ninth grade affects students' future occupational expectations in $11^{th}$ grade. I classify occupations by their math content following O*NET guidelines. I find that both men and women with higher math test scores and self-efficacy are more likely to expect a future occupation with higher math-related content. These results suggest that the preferences for major choices may be a function of early-life skills.

women would in fact benefit from enrolling and graduating from math-intensive fields is scarce. As a result, I present evidence on the returns to STEM majors for college students.

## 4.6 Labor Market Outcomes

**Returns to College Majors: Conceptual Framework**

While STEM-promoting policies may create important non-pecuniary benefits (Anaya et al., 2017), understanding the wage returns associated with these majors is a first-order concern for quantifying the benefits arising from such interventions. An extensive literature has estimated the returns to graduating from different majors. Altonji et al. (2012) highlight papers which have previously estimated gender-specific returns. A common empirical strategy, followed by Altonji (1993), Rumberger and Thomas (1993), and Eide (1994), among others, estimates a linear regression with controls for pre-college test scores. These papers find positive returns for women graduating from engineering, math and science degrees, relative to a degree in education. Altonji et al. (2016) report similar findings using ACS data without controls for test scores. While these results indicate positive returns to STEM degrees, they do not account for sorting into majors on unobserved characteristics, potentially resulting in biased estimates of the returns to major.[40]

[40] Jiang (2018) advances this literature by introducing a discrete choice model of college majors, where she finds positive returns to STEM-completion for women, relative to non-STEM fields. However, classifying majors in a binary fashion takes away from the analysis of heterogeneous returns across major pairings, as, for instance, the returns to STEM may vary depending on the major used as the counterfactual. Moreover, I discuss below how the returns to major completion differ from the returns to enrollment, a critical consideration in a sequential major choice model. Humphries et al. (2017) address these issues using Swedish data, but they do not examine gender-specific returns.

Furthermore, in a context of sequential nature of major choices, the returns to major completion capture a different parameter than the returns to enrollment, as the latter incorporate the possibility that a student may not subsequently complete the major. In fact, when considering the benefits arising from a policy nudging students to enroll in a different major, the policymaker should be interested in the latter parameter, which represents a linear combination of the wages of major completers, major switchers and college dropouts. Using the Quandt (1958) switching regression framework, I define the wages associated with any initial major $m_1$ as:

$$W_{m_1} = D_{m_1,G}W_{m_1,G} + D_{m_1,S}W_{m_1,S} + D_{m_1,D}W_{m_1,D} \ \ \forall \, m_1 \ \in \mathcal{M}_1 \tag{4.12}$$

where $D_{m_1,G}$, $D_{m_1,S}$, and $D_{m_1,D}$ represent dummy variables for individuals graduating from field $m_1$, switching to a different major, or dropping out of college, respectively. $W_{m_1,k}$ is the hourly wage associated with each of these outcomes. Letting $E[.]$ denote the expected value taken with respect to the distribution of $(\boldsymbol{X},\boldsymbol{\theta})$, I define the returns to enrollment in major $m_1$ as follows:

$$ATE_{m_1,m_k} = E\left[W_{m_1} - W_{m_k}\right] \ \ \forall \, m_k \ \in \mathcal{M}_1 \tag{4.13}$$

As the discrete choice model presented in section 4.3 allows me to recover the latent wages across initial majors, I can estimate the gender-specific average returns to STEM enrollment from equation (4.13) using model estimates. Heckman et al. (2016) show that the difference in the average returns and the observed wage difference across any two majors is explained both by selection bias, defined by the difference in latent wages in major $m_k$ for those in this major against those in STEM-fields, and by the sorting gains parameter. This parameter captures the possibility that students who have the most to gain from STEM majors may be the ones enrolled in these majors.[41] The difference in observed wages and the average returns to enrollment in major $m_1$ relative to major $m_k$ is given by:

$$
\begin{aligned}
\underbrace{E[Y_E|E=M_1]-[Y_E|E=M_k]}_{\text{Observed Difference}} = {} & \underbrace{E[Y_{E=M_1}-Y_{E=M_k}]}_{\text{ATE}\quad\text{(Enrollment)}}+ \\
& \underbrace{E[Y_{E=M_1}-Y_{E=M_k}|E=M_1]-E[Y_{E=M_1}-Y_{E=M_k}]}_{\text{Sorting Gains}}+ \\
& \underbrace{E[Y_{E=M_2}|E=M_1]-E[Y_{E=M_k}|E=M_k]}_{\text{Selection Bias}} \qquad (4.14)
\end{aligned}
$$

In Appendix E.1, I show the contribution of selection bias and sorting-on-gains for explaining the differences in the estimated wage benefits.

**Returns to College Majors: Empirical Evidence**

In the first panel of Figure 4.8, I present the average returns to STEM-enrollment for women relative to various alternative majors and compare it to observed wage differences across initial major pairs. First, as shown in Table C.1.4, the wages of women who start in STEM exced larger than those in the life sciences, other majors and non-declared students by upwards of 10 percent. Nonetheless, the average returns to STEM enrollment, estimated using simulated parameters from

[41]In the context of the latent wage equation (4.5), the sorting gains parameter would differ from zero if $\alpha_{m_1} \neq \alpha_{m_k}$. Selection bias and sorting gains can be explained by agents sorting-into-majors based both on observed ($\boldsymbol{X}$) and unobserved characteristics ($\boldsymbol{\theta}$).

the model, are lower than raw wage differences. The returns to STEM enrollment for women are heterogeneous across major pairings, as starting in STEM instead of in the life sciences yields an expected wage gain of 10 percent, but a negative return of 3 percent against business majors. Meanwhile, the returns to STEM relative to the life sciences, a major in the 'Other' category and for not declaring a major reach 5-8%, but the returns relative to health-related fields reach close to negative 20%.[42] In the second panel, I present the average returns to STEM completion, where I find positive returns relative to graduating with a life science or majors in the 'Other' category, as well as against college dropouts. The returns to enrollment are lower than those at graduation as the former parameter captures the possibility of subsequent dropout or switching into lower paying fields. As a result, for the returns to graduation to represent a policy-relevant parameter, students would need to be able to directly choose their major at college graduation. This is not an actionable margin in a model with sequential major choices.

In Figure C.1.4, I present the returns to enrollment in STEM for men, which are different than for women. Enrolling in STEM delivers large positive returns relative to any other field, except for in business, where the returns are indistinguishable from zero. The average treatment effect associated with STEM exceeds 20 percent vis-a-vis starting in the life sciences, 'Other' majors or not declaring a major. Why are the returns to enrolling in STEM significantly for women than they are for men? Differential STEM completion rates could explain part of the effect, as

[42]In Appendix E.1, I estimate the returns to STEM enrollment using various reduced-form approaches, including OLS, OLS with test scores as control variables and nearest-neighbor matching techniques, and show these estimates are significantly different than the average returns defined in equation (4.13).

just 44 percent of women finish this degree compared to 61 percent of men. On the other hand, following equations (4.5) and (4.12), I can examine gender differences in potential wages of STEM graduates. The comparison of potential wages corrects for endogenous sorting into completion, as these wages represent expected outcomes for any STEM enrollee if he/she were to graduate. As I find that latent wages for male STEM graduates exceed those of women by 15 percent, these results indicate that wage discrimination still plays a significant role in these fields.[43] While these results do not control for post-graduation occupational sorting, Goldin (2014) has found significant gender differences in within-STEM-occupation wages, indicating a sizable margin for pay disparities.

**Heterogeneous Returns to Majors**

The average returns to enrollment are computed by integrating out the latent skill distribution, yet may be heterogeneous across the ability vector, depending on the returns to each component of skills in both leading to college graduation and in increasing labor market productivity. I examine how the average treatment effect of enrolling in major $m_1$ varies across the unobserved ability distribution in:

$$ATE_{m_1,m_k}(\theta_C = \underline{\theta}, \theta_{SE} = \overline{\theta}) = E\left[W_{m_1} - W_{m_k} | \theta_C = \underline{\theta}, \theta_{SE} = \overline{\theta}\right] \quad \forall \, m_k \, \in \mathcal{M}_1$$

Furthermore, the returns to major $m_1$ ($ATE_{m_1,m_k}$) may also differ across students who chose to enroll in major $m_1$, given by the treatment on the treated (TT) parameter, and those who instead enrolled in major $m_k$, captured by the treatment

[43] The gender gap in potential wages in STEM equals 14 percent after controlling for gender differences in latent math skills.

on the untreated parameter (TUT). These parameters are defined as follows:

$$TT_{m_1,m_j} = E\left[W_{m_1} - W_{m_j} | D_{m_1} = 1\right] \quad \forall\, m_j \ \in \mathcal{M}_1 \tag{4.15}$$

$$TUT_{m_1,m_j} = E\left[W_{m_1} - W_{m_j} | D_{m_j} = 1\right] \quad \forall\, m_j \ \in \mathcal{M}_1 \tag{4.16}$$

In Table 4.5, I present the estimated returns to STEM enrollment. The returns from enrolling in STEM for women who have done so (TT) are positive relative to other majors (except for health fields), and larger than the estimated ATE across all alternative choices. As a result, students who stand to benefit the most from enrolling in STEM are more likely to have done so. On the other hand, the TUT parameters are negative across all major choices, indicating that had women enrolled in these majors instead started in a STEM field, they would have earned lower wages. The difference between the TT and TUT is statistically significant across all majors, which provides further confirmation of sorting into STEM (Heckman et al. 2016).

I examine heterogeneous returns by math ability in Panels B and C. For women below the median in each component of math ability, the average returns from enrolling in STEM are significantly smaller than the average treatment effect against all other majors, with positive returns relative to the life sciences, other degrees and not-declaring an initial major. The returns for those above the median of each component are significantly larger, yet not statistically different from zero in business fields, and remaining negative for health majors. On the other hand, the estimated returns are larger for women in the top decile of each component of math ability across most alternatives. For instance, I find that women in the top problem solving decile would enjoy wage returns of 16 percent by enrolling in STEM instead of in the life sciences and 10 percent by choosing STEM instead of not declaring a major. I find similar-sized returns for women in the top self-efficacy decile. The heterogeneity in the estimated returns suggests that while the average returns associated with a math-intensive field are lower than the observed wage differences across majors, high-skilled women would generally benefit from starting in STEM. While the results presented so far indicate that broad-based programs aimed at increasing women's STEM enrollment rates would lead to limited improvements in early-career labor market outcomes, increased female participation in STEM may have significant non-pecuniary benefits. For instance through parental role modeling effects for future generations. Furthermore, these results indicate that targeted programs aimed at high math ability women would uniformly yield positive returns.

**Conditional Returns to STEM Completion**

In a sequential model of major choices, students are not able to directly choose their major at college graduation. However, as they decide whether to complete their initial major after enrollment, understanding whether completion would deliver positive returns relative to switching majors or dropping out can inform students whether they should persist in their initial choice. The conditional returns to major completion are defined as follows:

$$ATE_{m_1,S} = E\left[W_{m_1,G} - W_{m_1,S} | D_{m_1} = 1\right] \quad \forall\, m_1 \in \mathcal{M}_1 \tag{4.17}$$

$$ATE_{m_1,D} = E\left[W_{m_1,G} - W_{m_1,D} | D_{m_1} = 1\right] \quad \forall\, m_1 \in \mathcal{M}_1 \tag{4.18}$$

The average treatment effect parameters defined above represent the benefits from completing major $m_1$ after having enrolled in it, relative to switching to a different field (equation (4.17)) or dropping out (equation (4.18)).[44] I present the estimated conditional returns to STEM completion in Table 4.6. The observed differences for students across their final decision (first row) do not represent a causal estimate of the returns to STEM completion. The second row estimates a regression including test scores, which indicate that women who graduate from STEM enjoy positive returns relative to switchers and college dropouts. As in equation (4.14), these results do not represent causal estimates of the returns to graduation. I present the causal returns following the discrete choice model in the second panel, where I find the average treatment effect for STEM completion for women is positive and significant both relative to switching majors (6.7 percent) and to dropping out from college (34.2 percent). For men, meanwhile, I both treatment effect margins are large and significant, exceeding 35 percent.

The difference in the treatment effect parameters shows significant sorting at STEM completion, as well. For instance, the treatment on the treated returns equal 15 percent for women graduating from STEM relative to major-switchers, which is significantly larger than the ATE presented above. On the other hand, the TUT parameter is small and not different from zero.

I find similar results in Panels B and C, where I analyze heterogeneous returns across both dimensions of math ability. The ATE associated with STEM completion,

[44]These returns are defined for agents who had initially enrolled in major $m_1$, though the parameters could be estimated for individuals in any other major. However, as noted by Heckman et al. (2018), these returns would not correspond to an actionable decision for an agent, so I restrict my attention to parameters which hold potential policy relevance.

relative to switching majors, is negative for women below the median of the problem solving distribution, while exceeding 11 percent for those above the median and 17 percent for those in the top decile. This result highlights the productivity of math ability in STEM-related fields, as higher skilled women earn higher wages by completing these degrees. I find similar results in the self-efficacy component, although the heterogeneity is less pronounced, as women in the top decile earn an average return of 8.7 percent. These results indicate that well-targeted policies aimed at increasing STEM completion among female enrollees may lead result in higher hourly wages. Given the heterogeneous returns to STEM majors by pre-college math ability and the sorting on math skills found in Section 4.4, I next examine whether skill-based interventions can offer a pathway for increasing women's STEM participation along with early-career labor market outcomes.

## 4.7 Policy Simulation: Math Self-Efficacy Intervention

Colleges across the country have implemented policies aimed at boosting students' STEM participation and subsequent completion rates, ranging from mentoring initiatives, STEM-program exposure, increased lab experience and summer preparation programs.[45] In this section, I follow the LATE framework, introduced by Imbens and Angrist (1994), to capture the effect of these interventions on any outcome

[45] These policies are summarized in EOP (2014). Cal Poly Pomona has launched a program which combines faculty mentoring, role models and an orientation for STEM enrollees. Alma College has a program focused on first-year STEM students, which offers access to various research opportunities. Mary Baldiwn College has an initiative offering summer research support and faculty mentorship for women in STEM fields. Michigan State University has launched a program to support under-prepared STEM students prior to matriculation by offering targeted courses and access to STEM faculty.

variable of interest $Y$. This framework allows me to separate the impact of STEM-promoting policies on students affected by the intervention (compliers) as well as those unaffected (always-takers and never-takers).[46] The effect of any policy $p'$ on outcome $Y$ is given by:

$$
\begin{aligned}
&\Delta^Y = E[Y(p') - Y] = \\
&\quad E[Y(p') - Y|D_s(p') = 1, D_s = 0] \times \underbrace{P[D_s(p') = 1, D_s = 0]}_{STEM\ Enrollment\ Compliers} + \\
&\quad E[Y(p') - Y|D_s(p') = 1, D_s = 1] \times \underbrace{P[D_s(p') = 1, D_s = 1]}_{STEM\ Enrollment\ Always-Takers} + \\
&\quad E[Y(p') - Y|D_s(p') = 0, D_s = 0] \times \underbrace{P[D_s(p') = 0, D_s = 0]}_{STEM\ Enrollment\ Never-Takers}
\end{aligned} \tag{4.19}
$$

where $D_s$ is a dummy variable which equals one for students enrolled in STEM. Equation (4.19) indicates that the aggregate effect of policy $p'$ on outcome variable $Y$ can be estimated by the linear combination of the effect on STEM always-takers, never-takers, and compliers, who are the students changing the enrollment decision due to the policy.[47] The effect of these interventions may vary by students' latent ability, depending on the nature of the policy and the outcome variable of interest. I focus my attention on evaluating the potential benefits arising from skill-based

[46]Different types of policies may have impacts on students not directly changing majors as a consequence of the intervention. For instance, a mentoring-based policy may successfully lead students to switch into STEM (compliers), while also increasing the likelihood of college graduation ($Y_\tau$) for students not directly switching their initial major due to the policy. The LATE framework allows me to capture the effect that different policies may have on a variety of relevant outcome variables, even for students not changing their enrollment decision.

[47]Throughout this section, I define response types by agents' initial major decisions. As a result, always-takers represent students who enroll in STEM both under baseline as well as in policy $p'$. Never-takers are those who do not enroll in STEM in either case. Compliers are those who choose to enroll in STEM as a function of $p'$, yet had not done so in the baseline. I test for the presence of defiers in the context of the simulated policies.

interventions for female college enrollees.[48]

### Math Self-Efficacy Based Interventions

The results presented in Sections 4.5 and 4.6 indicate that boosting math problem solving ability and self-efficacy would result in increased female participation in STEM. However, although cognitive and non-cognitive skills are highly malleable in the early years of life, non-cognitive skills are malleable through adolescence, unlike cognitive sills (Kautz et al., 2014). As a result, policies aimed at boosting women's math self-efficacy in high school could have greater effectiveness than those focused on the problem solving component. In this context, Huang (2013) has found that gender gaps in math self-efficacy expand from 0.06 SDs from middle school to 0.20 SDs in early high school. The psychology literature has found different strategies to be successful at increasing self-efficacy. Siegle and McCoach (2007) found a four-week course focused on improving high school math teachers' self-efficacy instructional strategies, which encompassed improving teacher feedback, establishing goals and presenting models of success, boosted students' math self-efficacy by 0.46 standard deviations. Cordero et al. (2010) and Betz and Schifano (2000) have similarly found positive effects of student-level self-efficacy interventions.

Following this literature, I use the estimated model parameters to examine the impact of an increase in women's math self-efficacy on STEM participation rates and on early-career labor market outcomes. As the psychology literature is not precise about the feasibility of interventions of varying magnitudes, I examine the

[48]In Appendix F.1, I examine the effects of a "nudging" policy, which would target female students closest to having started in STEM but who chose not to do so. These students are identified with the estimated utility parameters associated with each major (equations (4.1) and (4.3)), thus creating a cardinal ranking of all major choices for each student.

effects of simulated policies delivering math self-efficacy increases ranging from 0.1 to 1 standard deviations.[49] Despite the positive correlation between $\theta_{SE}$ and $\theta_C$, I assume that self-efficacy interventions would not jointly affect women's problem solving ability. As a result, the estimated impacts presented below likely represent a lower bound on the potential effect arising from self-efficacy-based policies.

**Effect on STEM Enrollment Rates**

I first examine the effect of a self-efficacy boost on STEM enrollment rates ($Y^E$). Following equation (4.19), the aggregate effect on $Y^E$ is fully captured by the share of compliers, as these students would be the sole group changing their decision on the basis of the simulated intervention $p'$. The effect on enrollment rates is thus given by:

$$\Delta^E = E[D_s(p') - D_s] = \underbrace{P[D_s(p') = 1, D_s = 0]}_{Compliers}$$

Increasing women's math self-efficacy would raise STEM enrollment rates, as measured by the share of compliers. An 0.5 SD boost in $\theta_{SE}$ would move enrollment rates from 4.5 percent to six percent, whereas an increase in a full standard deviation would further increase them to 7.7 percent, reducing the gender gap in STEM enrollment by almost one-fourth. This policy could have differential impacts across the $\theta_C$ distribution, depending on the complementarity of the two components of math ability in STEM fields. In the first panel of Figure 4.9, I examine how baseline

[49] I am agnostic as to the nature of the policy intervention which would deliver self-efficacy increases in the 0.1-1 SD range. The psychology literature indicates these interventions fall within a reasonable policy range.

participation rates change for women at each decile of the problem solving distribution. For women in the bottom $\theta_C$ decile, a 0.5 SD increase in self-efficacy would move enrollment rates to from 1.1 percent to just 1.8 percent. The largest impact appears for women in the top problem solving decile, whose STEM enrollment rates would increase from 8.8 percent to 11.4 percent, thus confirming the non-linear sorting patterns presented in Section 4.5.[50] As larger sized interventions have an positive linear impact on participation rates, the optimal self-efficacy intervention would depend on the structure of the cost function of achieving such gains.[51] As information on the cost function is not available, I focus on a simulated $\theta_{SE}$ boost of 0.25 standard deviations for the rest of the chapter.[52]

In the second panel of Figure 4.9, I show that a 0.25 SD self-efficacy boost would yield the largest increase STEM enrollment for women in the top problem solving decile, amounting to a 1.5 percentage point increase in enrollment rates, or almost one-sixth of baseline enrollment rates. These findings indicate that skill-based interventions could have larger impacts if well-targeted to high-achieving women. As noted above, depending on how preferences are formed during childhood, an early-life math self-efficacy intervention could have larger impacts on STEM par-

[50]Policymakers could alternatively be interested in boosting STEM participation rates by 1 percentage point at one of the deciles of the problem solving distribution. Such an effort would require a boost of $\theta_{SE}$ of 0.2 SDs for women in the top problem solving decile, of 0.4 SDs for those in the middle decile and of 1.2 standard deviations for those in the bottom decile.

[51]The cost function could be convex in nature, with small $\theta_{SE}$ increases requiring small costs, like teacher-training programs, yet larger increases may require repeated interventions during childhood.

[52]Only 0.9% of female college enrollees would change their STEM enrollment decision under an intervention of this magnitude. The majority of the sample would be comprised by never-taker (94.7%). The rest of the sample represents women already enrolled in STEM who would remain in those fields (4.5%). The bulk of the aggregate effects on any other outcome variable $Y$ are thus explained by the impact of this policy on never-takers. There are no defiers in the sample.

ticipation if it shaped women's preferences to increase interest and enjoyment from studying these fields.

**Effect on STEM Graduation Rates**

While increasing enrollment rates is an important first step for gauging the effectiveness of any STEM-promoting policy, this effect may not translate into increased graduation rates. Unlike the effect at enrollment, these policies could have an additional impact on graduation rates through always- and never-takers, as long as the intervention increased the likelihood of STEM completion despite not having changed the enrollment margin. In the context of the self-efficacy intervention, this effect would take place through the productivity of math self-efficacy in leading to STEM completion.

I examine the effects of the intervention on STEM completion rates in Table 4.7. A 0.25 SD increase in math self-efficacy would increase the share of women graduating from this field from 2.9% to 3.6%, which represents a relative increase of almost 20 percent. This effect is largely driven by increased completion rates among compliers, which would move from 3 percent to 46 percent. The effect on women already enrolled in STEM (always-takers) plays an important role in the aggregate effect, as their graduation rates would increase from 43.8% to 47.6%, which fits in with the productivity of self-efficacy in STEM majors presented in Figure 4.7.

As with the effect on enrollment rates, there may be heterogeneous impacts of the simulated intervention across the math problem solving distribution. In the first panel of Figure 4.10, I present the effects on always-takers, whose graduation rates increase largely uniformly across all $\theta_C$ deciles. For compliers, there is significant heterogeneity in their completion rates after having switched into STEM. For instance, while only 13% of those in the bottom problem solving decile end up completing a math-intensive degree, this share is four times larger for those in the top decile, exceeding 60 percent. Since math problem solving ability is a necessary component for successfully completing a STEM degree, low-ability students who are nudged into STEM through the self-efficacy intervention would still be lacking a critical component for subsequent success in these fields. Finally, the effect on never-takers is largely zero, as self-efficacy is not productive in non-STEM majors. All in all, the simulated policy would have much larger effects on students in the top problem solving decile, upping their completion rates from 8.5% to 10.2%, with a corresponding increase of only 0.1 percentage points for those in the bottom decile. The results presented so far indicate that small interventions focused on non-cognitive skills can help in increasing women's participation and graduation from math-intensive fields, with larger impacts for high math achievers.

**Effect on Labor Market Outcomes**

While I have so far shown that small self-efficacy-based interventions can lead to increased STEM participation rates for women, an open question remains about whether this policy would lead to improved labor market outcomes, especially in the context of the heterogeneous returns to math-intensive majors shown in Section 4.6. A self-efficacy boost could affect labor market outcomes both through an increased likelihood of STEM completion but also through increased labor market productivity, given the positive returns to non-cognitive skills found by Heckman et al. (2006) and by Lindqvist and Vestman (2011).

In Table 4.8, I show that the simulated self-efficacy boost would increase hourly wages for female college enrollees by 0.4%. The aggregate impact follows from a linear combination of the effect across the three potential response groups, which may be heterogeneous depending on the productivity of self-efficacy and the returns to STEM relative to the baseline major choices of compliers. First, the $\theta_{SE}$ boost would result in a 0.35% increase in hourly wages among never-takers. The small return to this component of non-cognitive ability is explained by the fact that math self-efficacy does not increase productivity in non-STEM majors, as few students in this group end up graduating from math-intensive fields. The small aggregate effect of the simulated policy on hourly wages is thus explained by the impact on never-takers.[53]

For always-takers, on the other hand, the effect would be positive and significant, increasing hourly wages by almost 3 percent. In Figure 4.11, I explore heterogeneous wage effects of the simulated policy for women already enrolled in STEM. The simulated self-efficacy boost delivers sizable returns for STEM enrollees in the top $\theta_C$ decile, exceeding 4%, while remaining at around 1-1.5% for those below the median. The heterogeneity in these returns associated with the intervention is explained by the fact that both margins of math ability non-linearly increase the probability of STEM completion, but also through the direct productivity of both components of ability in the labor market. These results contribute to the nascent literature showing the productivity of non-cognitive skills in improving labor market

[53]There is no heterogeneity in these returns across the problem solving distribution. High math achieving women not in STEM would not enjoy further benefits from a $\theta_{SE}$ boost. The results are available upon request.

outcomes.

Lastly, I find that the impact on the women switching from other fields into STEM enrollment (compliers) is not be statistically different from zero. However, this result masks the differential impact for students switching away from different non-STEM majors, which may be an important source of heterogeneity, as the returns to STEM enrollment vary by the alternative major under consideration. Note that compliers represent the linear combination of students switching out of the set of non-STEM fields:

$$\underbrace{P[D_s(p')=1, D_s=0]}_{Compliers} = \underbrace{P[D_s(p')=1, D_{s1}=1]}_{Compliers:\,Life\ Sciences} + ... + \underbrace{P[D_s(p')=1, D_{s2}=1]}_{Compliers:\,Other}$$

I follow this framework to first examine the majors from which compliers are switching out of (Table 4.8). 45% of compliers switch out of 'Other' majors, similar to the non-STEM enrollment share in the full sample. The largest difference appears among students who had not declared a major, who constitute 30% of the complier sample, almost doubling the baseline share of female non-declarees. A sizable share of women currently not declaring a major at enrollment would instead choose to start in a math-intensive field with a small self-efficacy boost. I find that this intervention would have heterogeneous impacts depending on the alternative major under consideration.[54] Students moving out of the life sciences, Other majors, as well as non-declarees would enjoy positive returns, in the range of 5-9%. On the

[54]These results match the heterogeneous returns to STEM by alternative major option shown in Table 4.5. Nonetheless, the wage returns to compliers capture a different parameter than the average treatment effects presented in equation (4.13).

other hand, the benefits arising from switching-into-STEM for students in health and business majors would be largely negative, exceeding -20% and -3%, respectively. As a result, the aggregate null effect on compliers masks important heterogeneity for students depending on which major they are switching out of. Policymakers should thus consider the alternative major under consideration when assessing the labor market returns to STEM-promoting policies. In the second panel of Figure 4.11, I expand upon these results by analyzing the heterogeneous benefits from the $\theta_{SE}$ boost for compliers across the problem solving distribution. Echoing the results for always-takers, I find positive wage returns for high-achieving women, who would earn higher wages under the simulated policy, exceeding 7 percent for high math-ability compliers.

Building on the heterogeneous impacts found for always-takers and compliers, I analyze how the simulated policy would affect labor market outcomes for the full sample of female college enrollees across the math ability distribution in Figure 4.12. I once again find positive impacts for women above the median of the distribution, with hourly wages increasing by one percent for those in the top decile of the problem solving distribution. While the magnitude of the wage effect may not appear to be economically significant, the simulated self-efficacy intervention is small in magnitude. Larger labor market impacts could potentially be achieved with improved math non-cognitive skill development for girls prior to college entry. Moreover, the proposed intervention would have a sizable impact on the labor market outcomes of high-achieving women who would switch into STEM, thus indicating that in a context of time/budget constraints, identifying individuals who would benefit the most from such interventions is of paramount importance. Although the heterogeneous impacts presented so far depend on women's unobserved math problem solving component, policymakers could instead target the top test score performers. While test scores are a noisy measure of ability, I find similar, though slightly muted, wage effects for top female test score performers (Figure C.1.6). All in all, well-targeted skill development policies can go a long way towards increasing women's participation in STEM and in improving early-career labor market outcomes.

## 4.8 Conclusion

In recent years, women's under-representation in STEM has received increased attention in both the economics literature and in policy discussion. In this chapter, I have examined the interaction between pre-college ability and major choices, with the goal of understanding the factors driving women's participation in STEM majors. Building on the other two chapters presented in this dissertation, I have introduced a discrete choice model, aimed at better understanding how students choose their college majors. Moreover, I have identified a particular element of the non-cognitive skill vector, mathematical self-efficacy, which plays a crucial role in determining students' enrollment and graduation from STEM majors. The identification of a particular non-cognitive skill component in this chapter differs from the general conception of non-cognitive skills pursued in Chapter 2. Relative to Chapter 3, this chapter considers the endogeneity of college major choices, rather than painting a descriptive picture. As a result, in this chapter, I have presented estimates of the causal returns to college majors to differentiate between observed and causal earnings differences across college majors.

Relative to the existing literature, while previous work has examined the role of pre-college preparation in explaining gender gaps in college majors, this analysis has been largely based on observed math test scores. To overcome this limitation, I have introduced a measurement system, where I have found that gender gaps in math test scores overstate differences in math problem solving by upwards of 40 percent. I have further identified an important non-cognitive component of math ability, self-efficacy, and found gender gaps in this dimension, as well. The correlation between these two components is lower for women than it is for men, indicating a relative lack of women at the top of the joint math skill distribution. This difference is particularly relevant to the analysis of STEM participation, as students sort into these majors non-linearly based on both dimensions of math ability. Furthermore, self-efficacy has a sizable effect in explaining female drop out from math-intensive fields, yet this pattern does not appear for men. The shortfall of high-achieving who are confident in their math skills thus reduces their participation in STEM majors. While I have offered preliminary evidence from alternative data sources on the origins of gender differences in the correlation in these components of math skills, future research should further explore this issue given the importance of math ability in driving STEM enrollment.

Given the focus on increasing women's STEM participation rates, I have also brought evidence to a relatively understudied aspect of the debate, which is whether women's labor market outcomes would improve from STEM enrollment. While increasing participation in these fields may bring important non-pecuniary benefits, I find significant heterogeneity in the wage returns in STEM, depending both on the alternative major under consideration but also across the math ability distribution. As I find large returns to STEM participation for high math ability women, STEM-promoting policies targeted towards high math-achievers would also deliver improved labor market outcomes.

Lastly, building on an extensive literature showing the malleability of non-cognitive skills through late adolescence, I have explored whether self-efficacy interventions could help in closing gender gaps in STEM. I have found that small skill development interventions could increase STEM enrollment and graduation rates by upwards of 15 percent, with larger impacts for high math performers. Furthermore, as the self-efficacy boost results in a small increase in hourly wages for top female math achievers, non-cognitive skill development interventions offer a promising pathway for future policy development.

## 4.9 Tables and Figures

**Table 4.1:** Descriptive Statistics

| | Women (1) | Men (2) | Difference (3) |
|---|---|---|---|
| Observable Characteristics | | | |
| White | 0.665 | 0.677 | -0.011 |
| Both Parents | 0.815 | 0.836 | -0.021* |
| Parental Education | 15.65 | 15.88 | -0.228*** |
| Log Family Income | 10.88 | 11.02 | -0.140** |
| Educational Attainment | | | |
| College Dropout | 0.204 | 0.210 | -0.007 |
| Bachelor's Degree | 0.59 | 0.637 | -0.039*** |
| Graduate School | 0.198 | 0.152 | 0.046*** |
| Test Scores | | | |
| $10^{th}$ Grade Math Exam | -0.086 | 0.175 | -0.261*** |
| $12^{th}$ Grade Math Exam | -0.113 | 0.211 | -0.325*** |
| $10^{th}$ Grade Math Self-Efficacy | -0.134 | 0.209 | -0.343*** |
| $12^{th}$ Grade Math Self-Efficacy | -0.115 | 0.183 | -0.299*** |
| Math GPA | 0.103 | -0.064 | 0.168*** |
| Math SAT/ACT | -0.131 | 0.137 | -0.294*** |
| English SAT/ACT | 0.011 | -0.014 | 0.024 |
| $10^{th}$ Grade Reading Exam | 0.021 | -0.026 | 0.047 |
| English GPA | 0.185 | -0.174 | 0.359*** |
| (Log) Hourly Wage | 2.823 | 2.918 | -0.095*** |
| Observations | 2,510 | 2,010 | |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. Students are required to have reported grades/scores for all the test scores presented above. All test score and GPA measures are normalized (0,1) for comparability. GPA measures represent an average over math/English courses taken in high school. Hourly wages are reported for employed college graduates who had not completed a graduate degree by 2012. Wages are reported as natural logarithms.

**Table 4.2:** Gender Participation by Major

| | Initial Major | | | Final Major | | |
|---|---|---|---|---|---|---|
| | Women | Men | Difference | Women | Men | Difference |
| Share in Option | (1) | (2) | (3) | (4) | (5) | (6) |
| Math-Intensive STEM | 0.045 | 0.179 | -0.134*** | 0.032 | 0.158 | -0.126*** |
| | (0.207) | (0.383) | (0.009) | (0.177) | (0.365) | (0.004) |
| Life Sciences | 0.076 | 0.072 | 0.005 | 0.072 | 0.070 | 0.002 |
| | (0.266) | (0.258) | (0.008) | (0.316) | (0.255) | (0.009) |
| Business | 0.116 | 0.181 | -0.065*** | 0.118 | 0.176 | -0.058*** |
| | (0.320) | (0.385) | (0.010) | (0.323) | (0.381) | (0.010) |
| Health | 0.135 | 0.036 | 0.099*** | 0.105 | 0.042 | 0.064*** |
| | (0.342) | (0.186) | (0.008) | (0.307) | (0.199) | (0.008) |
| Other | 0.470 | 0.340 | 0.130*** | 0.468 | 0.345 | 0.124*** |
| | (0.499) | (0.474) | (0.015) | (0.495) | (0.475) | (0.015) |
| Not Declared | 0.158 | 0.193 | -0.035*** | | | |
| | (0.365) | (0.394) | (0.011) | | | |
| Not Graduated | | | | 0.204 | 0.210 | -0.006 |
| | | | | (0.403) | (0.408) | (0.012) |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. Columns (3) and (6) present the gender difference across initial majors and final outcomes from a two-tailed t-test. Math-intensive STEM fields include degrees in engineering, engineering-related fields, computer science, mathematics, economics, statistics and physics. Life science degrees include majors in agriculture (and related sciences), natural resources and conservation, family science, biology and related fields and other science technologies. Business degrees includes degrees in business, management and marketing. The "Other" group includes the the following college majors: Architecture, Anthropology, Art, Art History, Communications, Criminal Justice, Education, English, History, International Relations, Journalism, Literature, Pre-Law, Political Science, Psychology, Social Work, and Sociology, among others. The Health group is largely composed of majors in Nursing, Pre-Med, Pre-Vet, Pharmacy, Health and Physical Therapy.

**Table 4.3:** Sorting into STEM Majors and Enrollment and Graduation

| | STEM Enrollment | | STEM Completion | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| | (1) | (2) | (3) | (4) |
| White | -0.015 | -0.005 | 0.177 | 0.091* |
| | (0.010) | (0.020) | (0.096) | (0.049) |
| Fam. Income | 0.004* | 0.005 | 0.008 | -0.024* |
| | (0.002) | (0.004) | (0.026) | (0.014) |
| Dad in Field | 0.017 | 0.061* | 0.074 | 0.024 |
| | (0.016) | (0.031) | (0.158) | (0.072) |
| Mom in Field | 0.017 | 0.022 | -0.113 | 0.032 |
| | (0.027) | (0.055) | (0.250) | (0.130) |
| Math Test | 0.015* | 0.082*** | 0.072*** | 0.096*** |
| | (0.006) | (0.012) | (0.065) | (0.031) |
| Self-Efficacy | 0.017*** | 0.057*** | 0.117*** | 0.026 |
| | (0.004) | (0.009) | (0.054) | (0.028) |
| English Test | -0.004 | -0.038*** | 0.055 | 0.003 |
| | (0.006) | (0.011) | (0.060) | (0.039) |
| Constant | -0.032 | 0.155* | 0.198 | 0.759 |
| | (0.038) | (0.078) | (0.298) | (0.156)*** |
| $N$ | 2510 | 2010 | 110 | 460 |
| $R^2$ | 0.018 | 0.066 | 0.173 | 0.065 |
| Baseline Share | 0.045 | 0.179 | 0.442 | 0.626 |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: The first two columns include all the males and females in the college enrollee sample. I estimate a linear probability model with a dummy variable for STEM enrollment as the outcome variable. The Math Test score and the English Test variables represent the $10^{th}$ grade exam test scores, normalized (0,1). The Self-Efficacy measure is also taken in $10^{th}$ grade. The results are robust to other test score measures. The last two columns examine STEM completion rates among *students initially enrolled in STEM.* I once again estimate a linear probability model with STEM completion as the outcome variable. The set of explanatory variables remains the same across the four columns.

**Table 4.4:** Descriptive Statistics: Latent Factors v. Baseline Test Scores

**Panel A.** Latent Factors

| Men | | | | | Women | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Correlation | | | | | Correlation | | |
| Factor | Mean (SD) | $\theta_C$ | $\theta_{SE}$ | $\theta_R$ | Factor | Mean (SD) | $\theta_C$ | $\theta_{SE}$ | $\theta_R$ |
| $\theta_C$ | 0.166 | 1 | | | $\theta_C$ | -0.001 | 1 | | |
| | (0.682) | | | | | (0.647) | | | |
| $\theta_{SE}$ | 0.151 | 0.561 | 1 | | $\theta_{SE}$ | -0.001 | 0.471 | 1 | |
| | (0.746) | | | | | (0.761) | | | |
| $\theta_R$ | -0.129 | 0.863 | 0.483 | 1 | $\theta_R$ | 0 | 0.848 | 0.412 | 1 |
| | (0.873) | | | | | (0.808) | | | |

**Panel B.** Baseline Test Scores

| Men | | | | | Women | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Correlation | | | | | Correlation | | |
| Measure | Mean (SD) | Math | SE | Read | Measure | Mean (SD) | Math | SE | Read |
| Math Exam | 0.261 | 1 | | | Math Exam | 0 | 1 | | |
| | (0.949) | | | | | (0.988) | | | |
| Self-Efficacy | 0.343 | 0.347 | 1 | | Self-Efficacy | -0.001 | 0.292 | 1 | |
| | (0.995) | | | | | (0.957) | | | |
| Reading Exam | -0.048 | 0.642 | 0.178 | 1 | Reading Exam | 0 | 0.647 | 0.092 | 1 |
| | (0.957) | | | | | (1.017) | | | |

Source: Educational Longitudinal Study of 2002.
Note: Table 4.4 displays the mean, standard deviation and correlation between the three ability components separately by gender. $\theta_C$ represents the problem solving factor, $\theta_{SE}$ is the math self-efficacy component and $\theta_R$ is the reading ability component. Results are simulated from the estimates of the model. The second panel displays the mean, standard deviation and correlation between the three baseline math and reading test scores as well as the baseline math self-efficacy measure.

**Table 4.5:** Estimated Returns to STEM Enrollment for Women

**Panel A.** Returns to Enrollment Relative to Other Majors

| | Life Sciences | Business | Health | Other | Not Declared |
|---|---|---|---|---|---|
| ATE | 0.097 | -0.041 | -0.192 | 0.087 | 0.059 |
| | (0.005)*** | (0.005)*** | (0.005)*** | (0.005)*** | (0.005)*** |
| TT | 0.152 | -0.032 | -0.209 | 0.108 | 0.046 |
| | (0.021)*** | (0.013)** | (0.022)*** | (0.021)*** | (0.014)*** |
| TUT | 0.116 | -0.041 | -0.196 | 0.098 | 0.062 |
| | (0.010)*** | (0.013)*** | (0.007)*** | (0.006)*** | (0.007)*** |
| MTE | 0.060 | -0.044 | -0.207 | 0.079 | 0.040 |
| | (0.019)*** | (0.021)** | (0.022)*** | (0.021)*** | (0.022)* |

***Panel B.** Heterogeneous Returns by Math Problem Solving Ability

| | Life Sciences | Business | Health | Other | Not Declared |
|---|---|---|---|---|---|
| ATE (Low $\theta_C$) | 0.070 | -0.075 | -0.180 | 0.055 | 0.040 |
| | (0.006)*** | (0.006)*** | (0.006)*** | (0.006)*** | (0.006)*** |
| ATE (High $\theta_C$) | 0.124 | -0.008 | -0.204 | 0.119 | 0.077 |
| | (0.006)*** | (0.006) | (0.006)*** | (0.006)*** | (0.006)*** |
| ATE (Top $\theta_C$ Decile) | 0.158 | 0.025 | -0.201 | 0.153 | 0.099 |
| | (0.014)*** | (0.014) | (0.014)*** | (0.014)*** | (0.014)*** |

**Panel C.** Heterogeneous Returns by Math Self-Efficacy

| | Life Sciences | Business | Health | Other | Not Declared |
|---|---|---|---|---|---|
| ATE (Low $\theta_{SE}$) | 0.071 | -0.071 | -0.204 | 0.060 | 0.038 |
| | (0.006)*** | (0.006)*** | (0.006)*** | (0.006)*** | (0.006)*** |
| ATE (High $\theta_{SE}$) | 0.123 | -0.012 | -0.180 | 0.115 | 0.079 |
| | (0.006)*** | (0.006)* | (0.006)*** | (0.006)*** | (0.006)*** |
| ATE (Top $\theta_{SE}$ Decile) | 0.148 | 0.040 | -0.162 | 0.146 | 0.107 |
| | (0.014)*** | (0.014)*** | (0.014)*** | (0.014)*** | (0.014)*** |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: Table 4.5 presents the returns to enrollment to women from the simulated discrete choice model. The returns are estimated separately against each alternative major choice, $m_1 \in \mathcal{M}_\infty$. The average treatment effect (ATE) refers to the parameter indicated in equation (4.13). The treatment on the treated parameter (TT) refers to the parameter indicated in equation (4.15). The treatment on the untreated parameter (TUT) refers to the parameter indicated in equation (4.16). In Panel B, the ATE (Low $\theta_C$) denotes the benefits for women below the math problem solving median, whereas the (High $\theta_C$) refers to those above the median. A similar definition is used in Panel C.

**Table 4.6:** Estimated Returns to STEM Completion

*__Panel A.__ Returns to Completion by Gender

| | Women | | Men | |
|---|---|---|---|---|
| | v. Switchers | v. Dropouts | v. Switchers | v. Dropouts |
| Reduced Form | | | | |
| No Controls | 0.175 | 0.401 | 0.442 | 0.401 |
| | (0.144) | (0.014)*** | (0.143)*** | (0.075)*** |
| Full Model | 0.131 | 0.257 | 0.442 | 0.357 |
| | (0.149) | (0.155) | (0.088)*** | (0.077)*** |
| Model Estimates | | | | |
| ATE | 0.067 | 0.342 | 0.442 | 0.349 |
| | (0.014)*** | (0.014)*** | (0.010)*** | (0.010)*** |
| TT | 0.149 | 0.439 | 0.450 | 0.349 |
| | (0.021)*** | (0.021)*** | (0.013)*** | (0.013)*** |
| TUT | 0.042 | 0.218 | 0.450 | 0.349 |
| | (0.024) | (0.029)*** | (0.021)*** | (0.023)*** |
| MTE | 0.270 | 0.286 | 0.367 | 0.338 |
| | (0.099)** | (0.128)** | (0.069)*** | (0.076)*** |

**Panel B.** Heterogeneous Returns by Math Problem Solving Ability

| | Women | | Men | |
|---|---|---|---|---|
| | v. Switchers | v. Dropouts | v. Switchers | v. Dropouts |
| ATE (Low $\theta_C$) | -0.063 | 0.236 | 0.446 | 0.370 |
| | (0.026)*** | (0.026)*** | (0.018)*** | (0.018)*** |
| ATE (High $\theta_C$) | 0.118 | 0.383 | 0.440 | 0.342 |
| | (0.016)*** | (0.016)*** | (0.012)*** | (0.012)*** |
| ATE (Top $\theta_C$ Decile) | 0.172 | 0.461 | 0.438 | 0.320 |
| | (0.030)*** | (0.030)*** | (0.022)*** | (0.022)*** |

**Panel C.** Heterogeneous Returns by Math Self-Efficacy

| | Women | | Men | |
|---|---|---|---|---|
| | v. Switchers | v. Dropouts | v. Switchers | v. Dropouts |
| ATE (Low $\theta_{SE}$) | 0.071 | 0.291 | 0.439 | 0.355 |
| | (0.027)*** | (0.027)*** | (0.018)*** | (0.018)*** |
| ATE (High $\theta_{SE}$) | 0.066 | 0.360 | 0.442 | 0.346 |
| | (0.016)*** | (0.016)*** | (0.012)*** | (0.012)*** |
| ATE (Top $\theta_{SE}$ Decile) | 0.087 | 0.365 | 0.436 | 0.352 |
| | (0.030)*** | (0.030)*** | (0.022)*** | (0.022)*** |

Source: Educational Longitudinal Study of 2002. Note: * p<0.10, ** p<0.05, *** p<0.01. Note: Table 4.6 presents the returns to graduation to math-intensive STEM majors for women and men. The reduced form estimates presented in the first two rows of Panel A follow from pairwise comparisons of STEM graduates to switchers and dropouts, respectively, among initial STEM enrollees. The first row includes individual and family background characteristics as controls. The second row adds baseline test scores as control variables.

**Table 4.7:** Aggregate Effects of Self-Efficacy Intervention on Graduation Rates and Hourly Wages

| | Graduation Rates | | Hourly Wages | |
|---|---|---|---|---|
| | Baseline | Post-Intervention | Baseline | Post-Intervention |
| | (1) | (2) | (3) | (4) |
| Full Sample | 0.029 | 0.036 | 2.770 | 2.774 |
| | | (0.001)*** | | |
| Always-Takers | 0.438 | 0.476 | 2.778 | 2.807 |
| | | (0.001)*** | | (0.019)* |
| Compliers | 0.032 | 0.462 | 2.817 | 2.821 |
| | | (0.002)*** | | (0.043) |
| Never-Takers | 0.010 | 0.011 | 2.769 | 2.772 |
| | | (0.000)*** | | (0.001)*** |

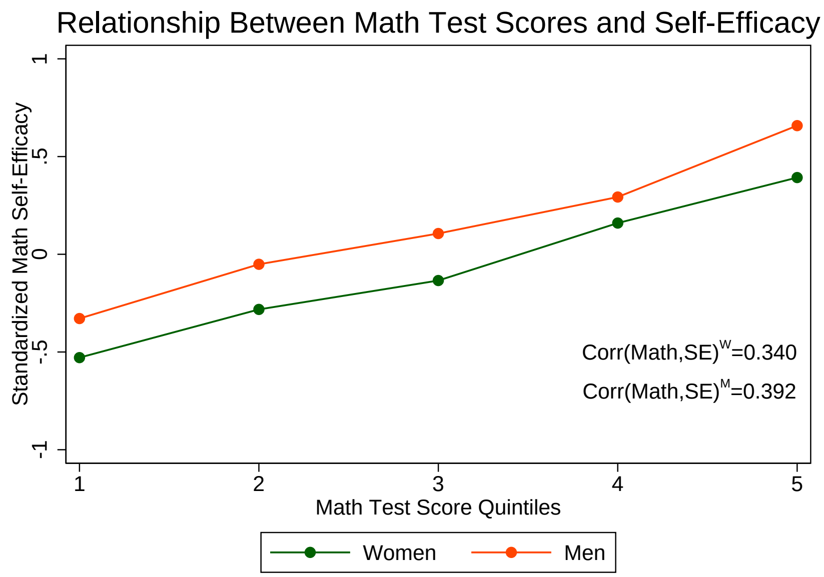
Source: Educational Longitudinal Study of 2002. Note: * p<0.10, ** p<0.05, *** p<0.01. Note: Table 4.7 examines the aggregate impact of the self-efficacy intervention on outcomes for female college students. The standard errors refer to a comparison of enrollment rates relative to the baseline. I explore the impact of the 0.25SD boost in $\theta_{SE}$ across the potential response groups and the SEs follow from a comparison of graduation rates relative to the baseline.

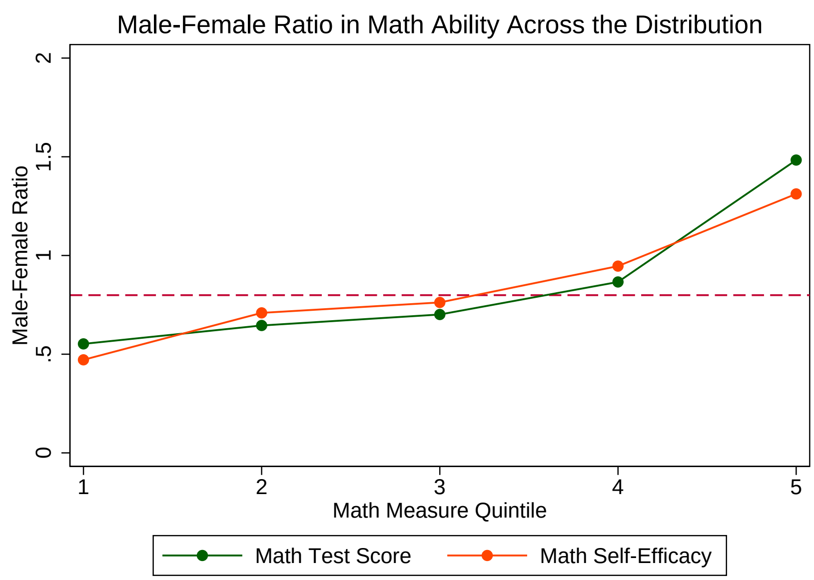
**Table 4.8:** Aggregate Effects of Self-Efficacy Intervention Across Complier Types

| | Composition | | Compliers' Hourly Wages | |
|---|---|---|---|---|
| | % in Sample | Complier Share | Baseline | Post-Intervention |
| | (1) | (2) | (3) | (4) |
| Life Sciences | 0.082 | 0.027 | 2.687 | 2.752 |
| | | | | (0.022)*** |
| Business | 0.123 | 0.092 | 2.833 | 2.799 |
| | | | | (0.018)* |
| Health | 0.142 | 0.126 | 2.949 | 2.684 |
| | | | | (0.016)*** |
| Other | 0.486 | 0.456 | 2.719 | 2.795 |
| | | | | (0.004)*** |
| Not Declared | 0.168 | 0.299 | 2.756 | 2.835 |
| | | | | (0.006)*** |

Source: Educational Longitudinal Study of 2002. Note: * p<0.10, ** p<0.05, *** p<0.01. Note: Table 4.7 examines the aggregate impact of the self-efficacy intervention on outcomes for female college students. This table explores the impact on hourly wages across these groups. Column (2) in Panel D compares the share of compliers in each of the five non-STEM majors to their baseline non-STEM participation in the full female sample. The fourth column compares their hourly wages post-intervention to their hourly wages in the baseline. The SEs correspond to a test of the difference of wages in the baseline v. post-intervention.

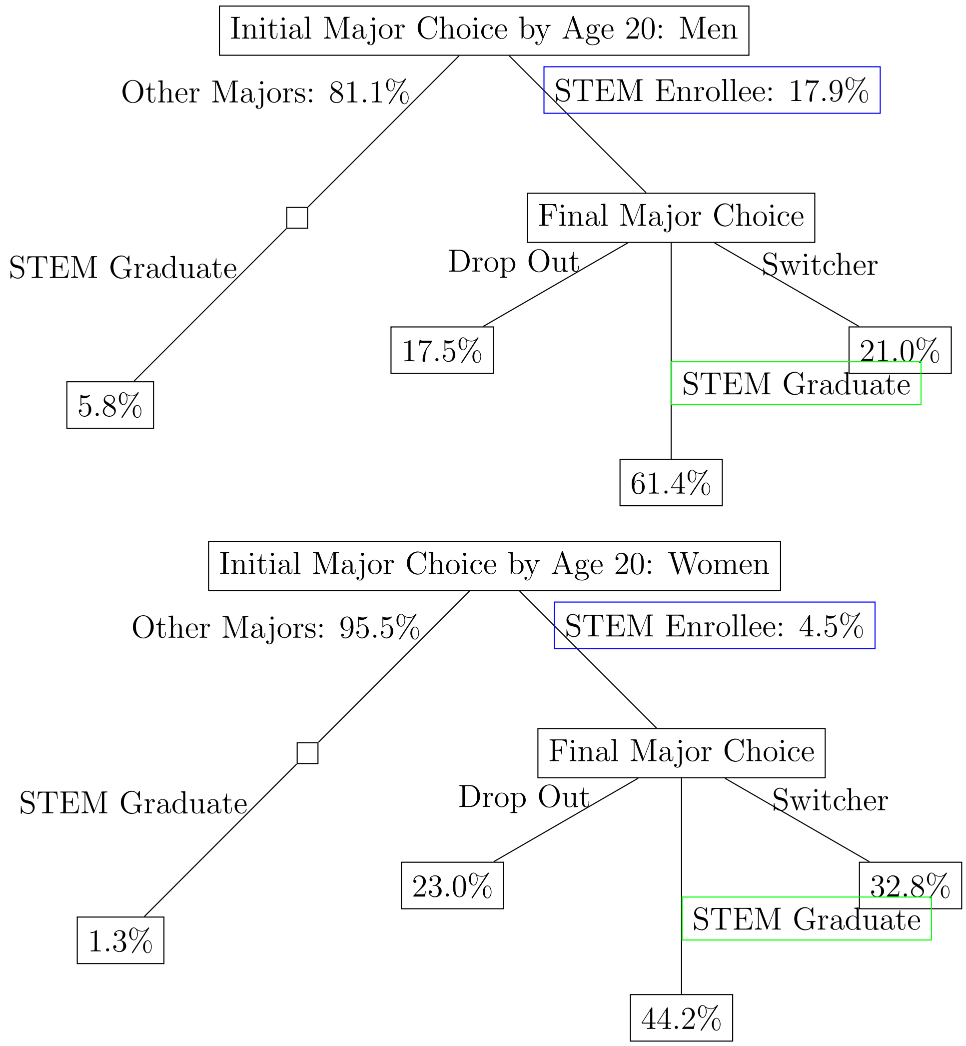
**Figure 4.1:** Math Test Score and Self-Efficacy Measures



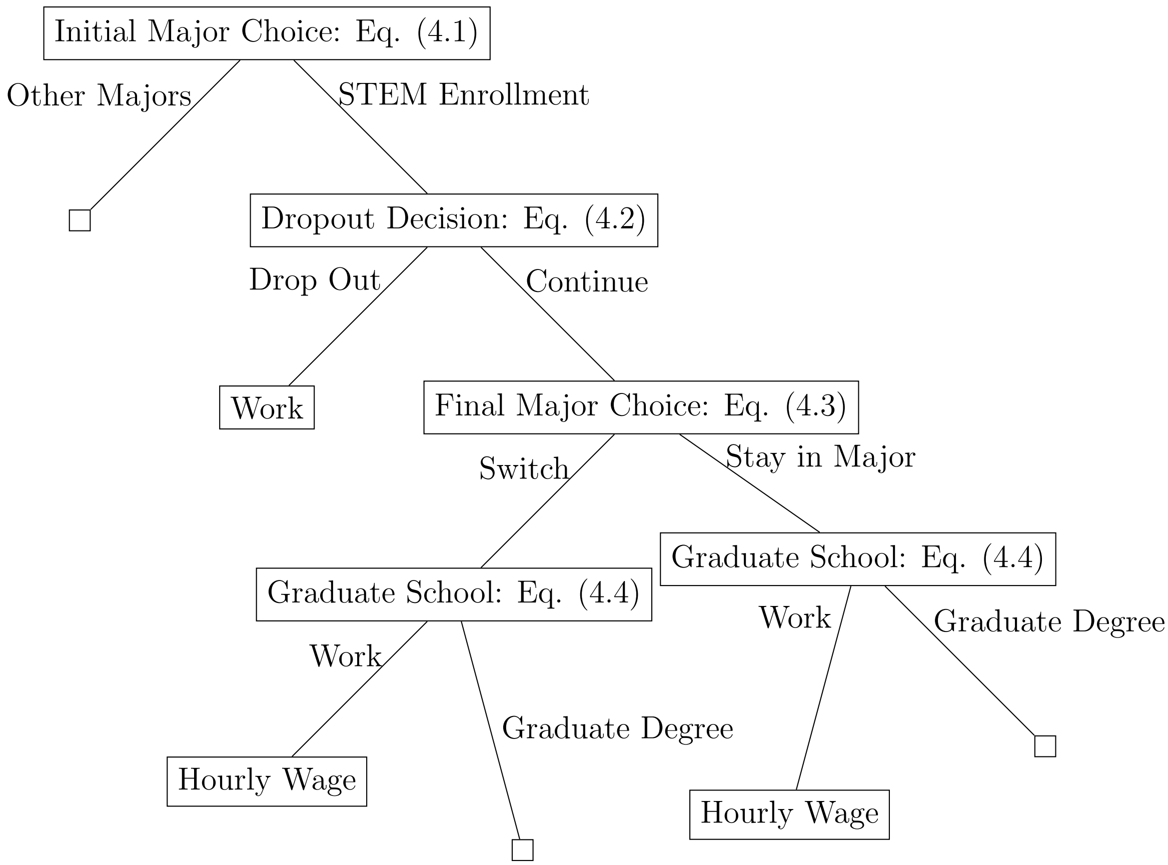

Source: Educational Longitudinal Study of 2002. Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. In Panel A, baseline self-efficacy scores are standardized normal (0,1), while math test scores are divided into quintiles. The second panel shows the ratio of men to women in each quintile of baseline math test scores and self-efficacy measures.

**Figure 4.2:** STEM Enrollment Patterns and Subsequent Completion by Gender



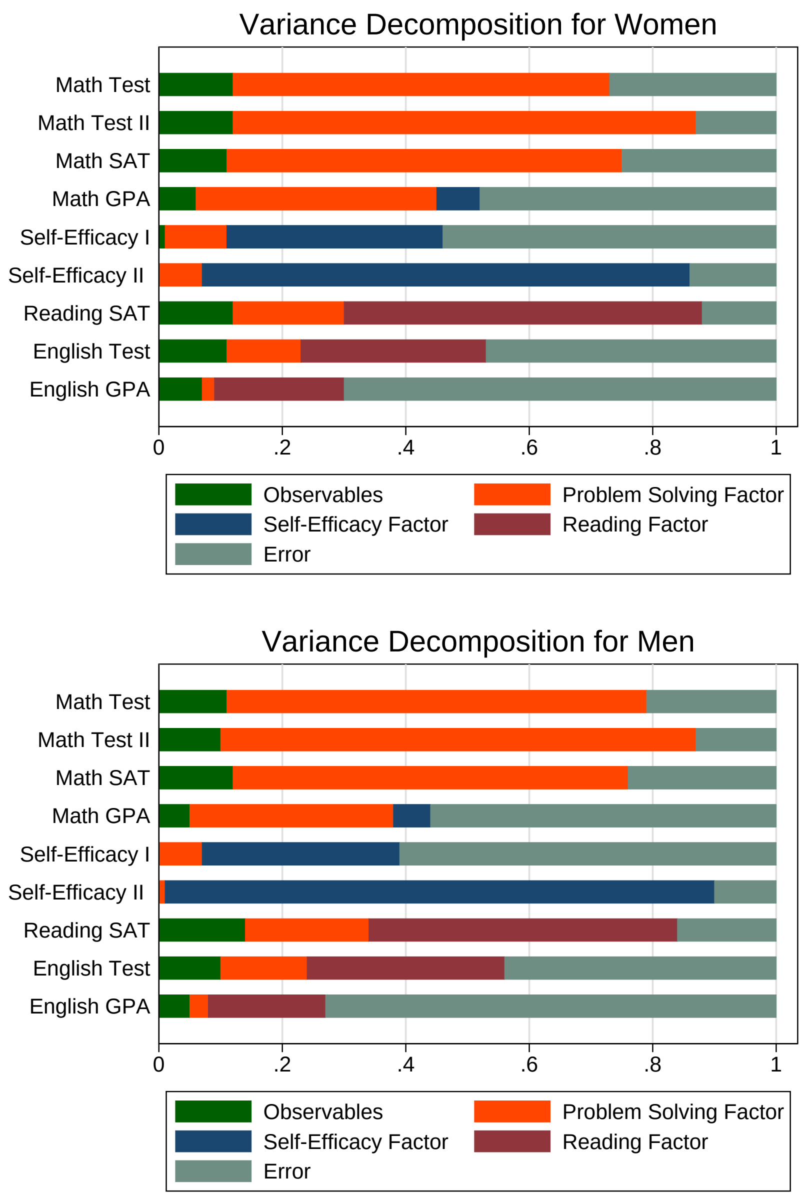

Source: Educational Longitudinal Study of 2002.
Note: This Figure depicts sequential progression through college majors for men and women by initial major choice (STEM and non-STEM). The left line depicts the share of students starting in other majors and the line below it, the share among these students who subsequently switch into a STEM major at graduation. The right line captures the share of men and women who start in a STEM field. In the second line, I then show the share of students in this group who end up completing a STEM degree, switching into another field, or dropping out of college altogether.

**Figure 4.3:** Structure of Discrete Choice Model



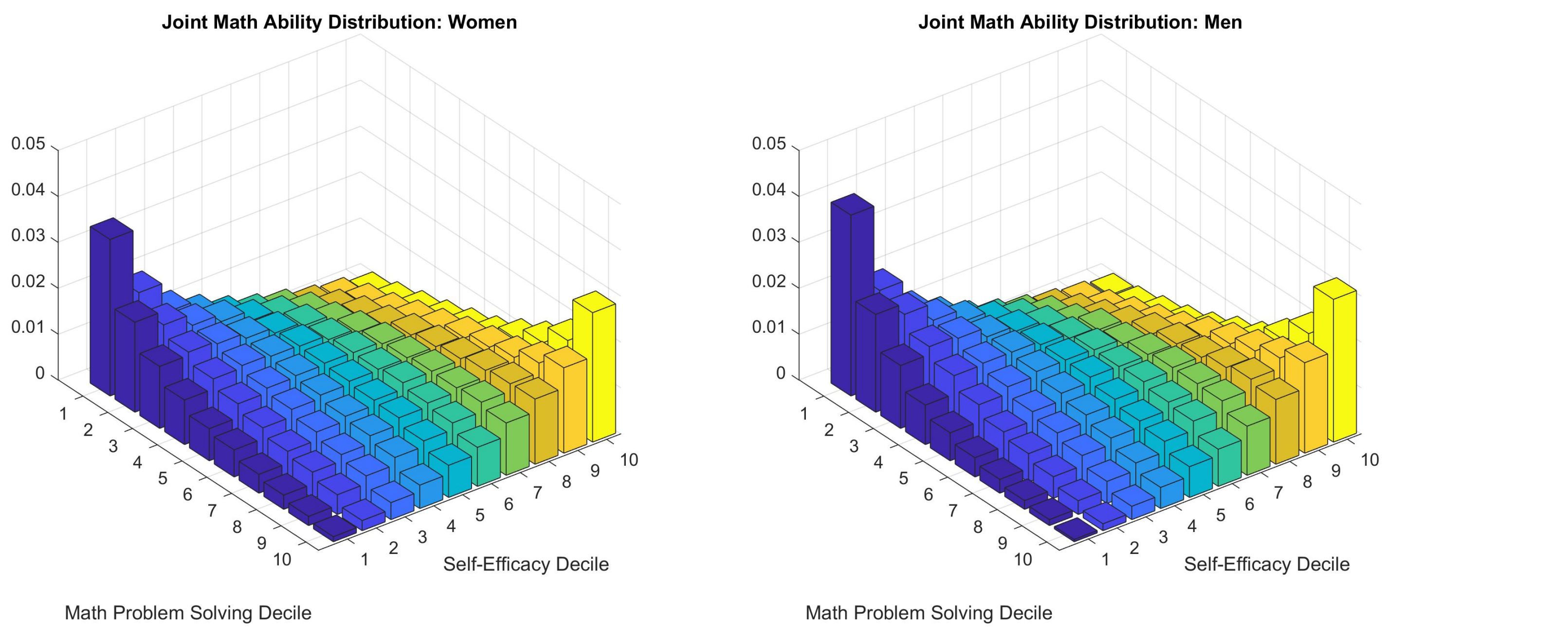

Source: Educational Longitudinal Study of 2002.
Note: This Figure depicts sequential progression through college majors for initial STEM enrollees. Despite this process not being depicted for students enrolled in 'Other' majors (due to limited space), college progression for these students, and for those in all other categories, follows the same pattern.

**Figure 4.4:** Measurement System: Variance Decomposition



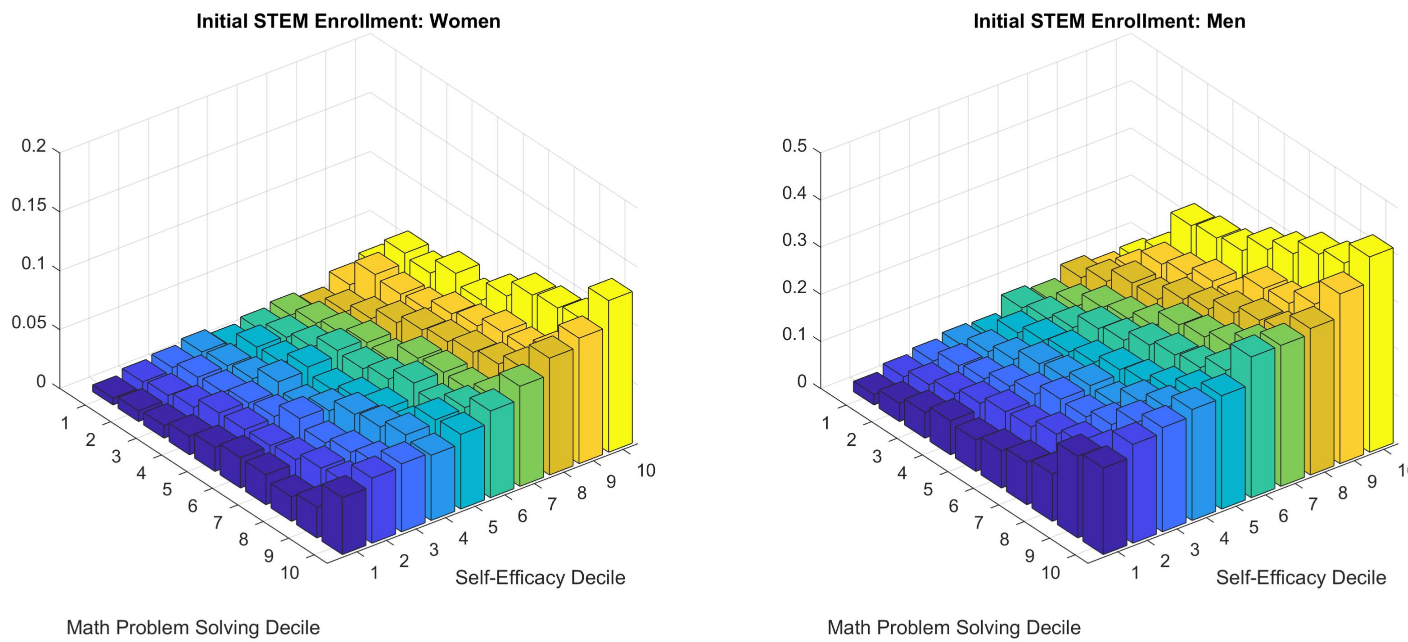

Source: Educational Longitudinal Study of 2002. Note: This Figure shows the contribution of each variable to the variance of test scores using the simulated sample the model. The row Observables indicates the share of the variance of the measurement variables explained by the observed variables: child's race, parental education and occupation, and household income. Each Factor bar indicates the share of the variance explained by each component of latent ability. Finally, the label Error term represents the share of each test score variance explained by the unobserved idiosyncratic error of the measurement system.

**Figure 4.5:** Joint Math Ability Density by Gender



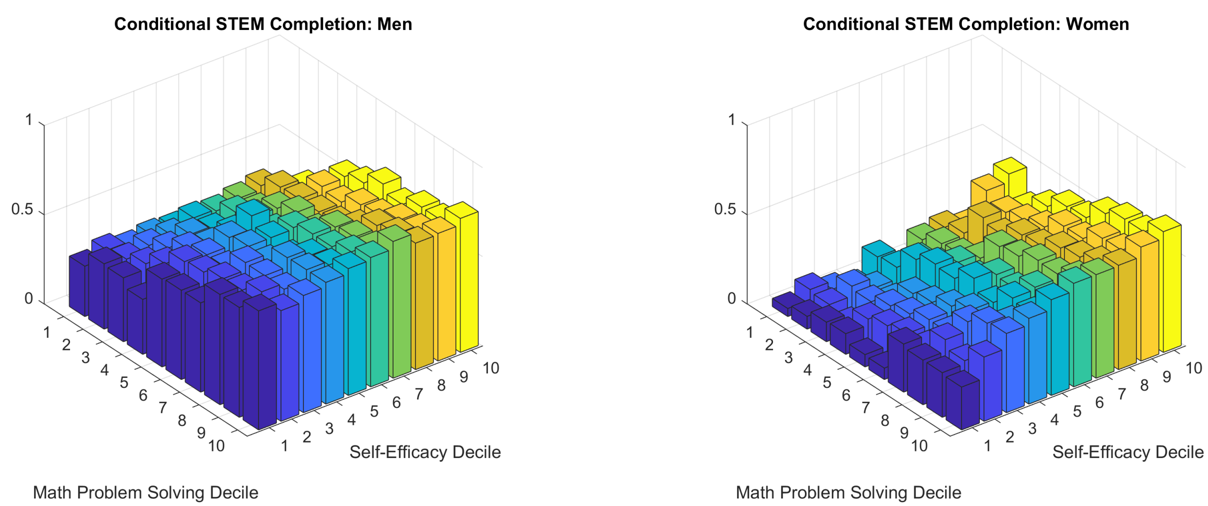

Source: Educational Longitudinal Study of 2002.
Note: This Figure shows the joint density of the math problem solving and the self-efficacy factors separately by gender. The density is presented as the share of individuals in each centile of the joint distribution.

**Figure 4.6:** Share Initially Enrolled in a STEM Field by Gender



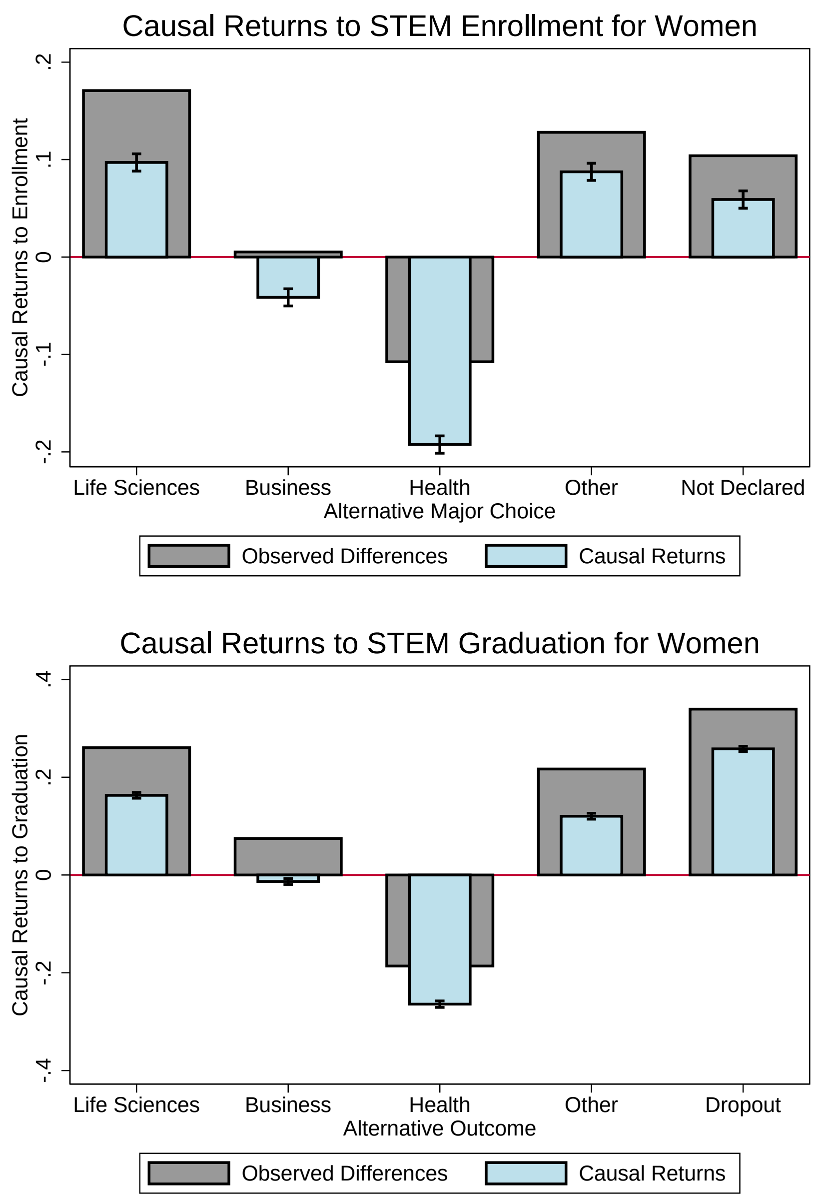

Source: Educational Longitudinal Study of 2002.
Note: Figure 4.6 shows the share of women and men initially enrolled in a math-intensive major in each joint decile of the math problem solving and the self-efficacy factors. The deciles of problem solving and self-efficacy are defined relative to the within-gender ability distribution. The results do not differ when defining ability deciles including both women and men.

**Figure 4.7:** Completion Rates for Students Initially Enrolled in STEM



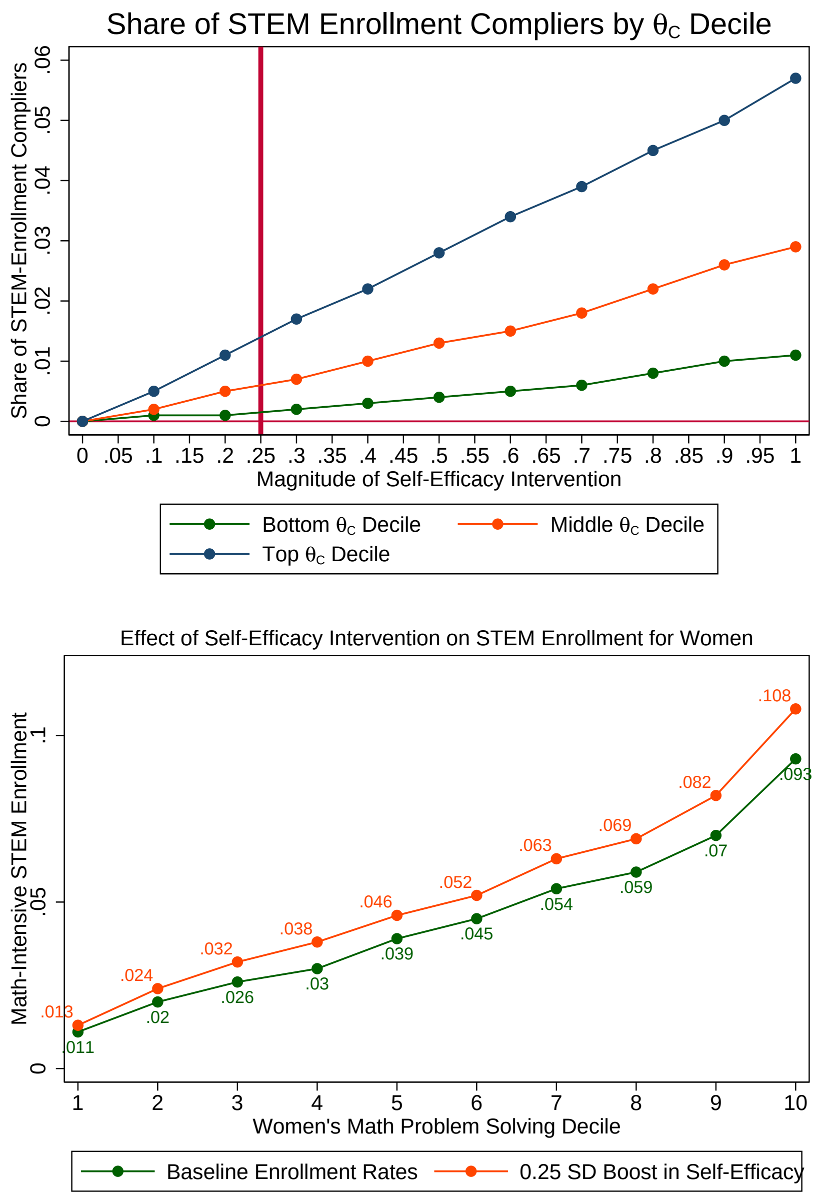

Source: Educational Longitudinal Study of 2002. Note: Figure 4.7 shows the share of women and men who graduate from a math-intensive major by age 26, after having started in one of these majors, by the joint decile of the math problem solving and the self-efficacy ability components. The deciles of problem solving and self-efficacy are defined relative to the within-gender ability distribution. The results do not differ when defining ability deciles including both women and men.

**Figure 4.8:** Estimated Causal Returns to STEM Majors for Women



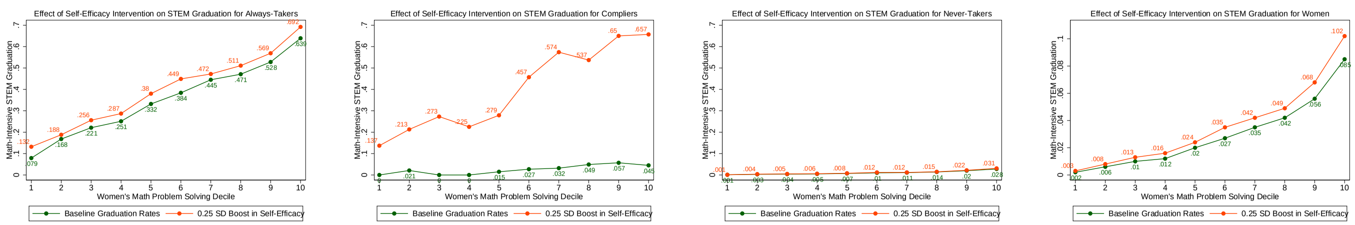

Source: Educational Longitudinal Study of 2002. Note: Figure 4.8 presents the returns to enrollment in math-intensive STEM majors. The returns are estimated separately against each alternative major choice, $m_1 \in \mathcal{M}_\infty$ and educational outcome $m_2 \in \mathcal{M}_\in$ in each panel, respectively. The returns presented represent the average treatment effect (ATE) of each major, as defined in equation (4.13). The returns to enrollment and graduation are compared to the raw wage differences among STEM enrollees and completers, respectively, against the alternative outcome. The 'Causal Returns' estimate follows from estimated parameters in the dynamic discrete choice model.

**Figure 4.9:** Estimated Impacts of Self-Efficacy Interventions on STEM Enrollment



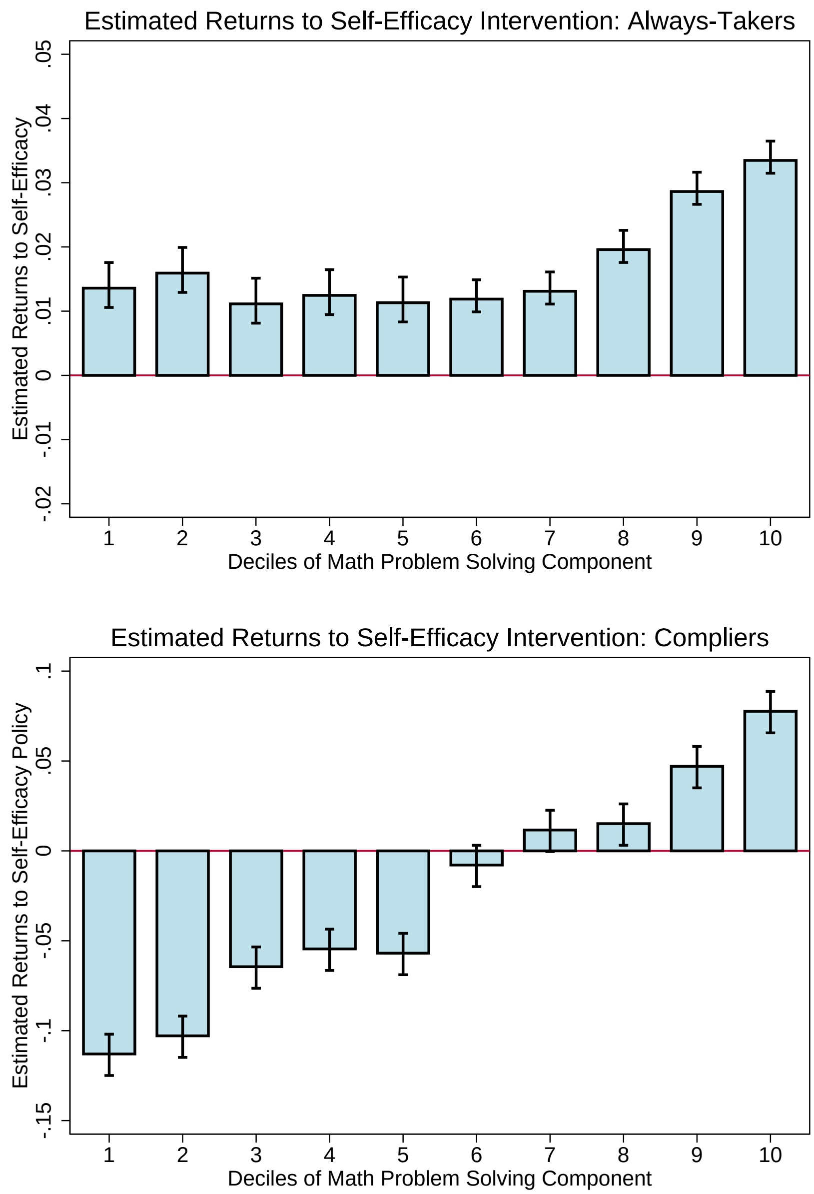

Source: Educational Longitudinal Study of 2002. Note: The first panel of Figure 4.9 examines the impacts of self-efficacy interventions of varying magnitudes (from 0.1 SD to 1 SD) on women's enrollment rates in STEM, as measured by the share of compliers at each decile of the $\theta_C$ distribution. The vertical red line is drawn at 0.25 SDs, which is the magnitude of the policy intervention analyzed in Section 4.7. The second panel analyzes the impact of the 0.25 SD intervention on aggregate enrollment rates across each decile of the $\theta_C$ distribution. The green line captures baseline enrollment rates, whereas the orange line analyzes enrollment rates after the intervention.

**Figure 4.10:** Effects of Self-Efficacy Intervention on STEM Graduation by Potential Response Group



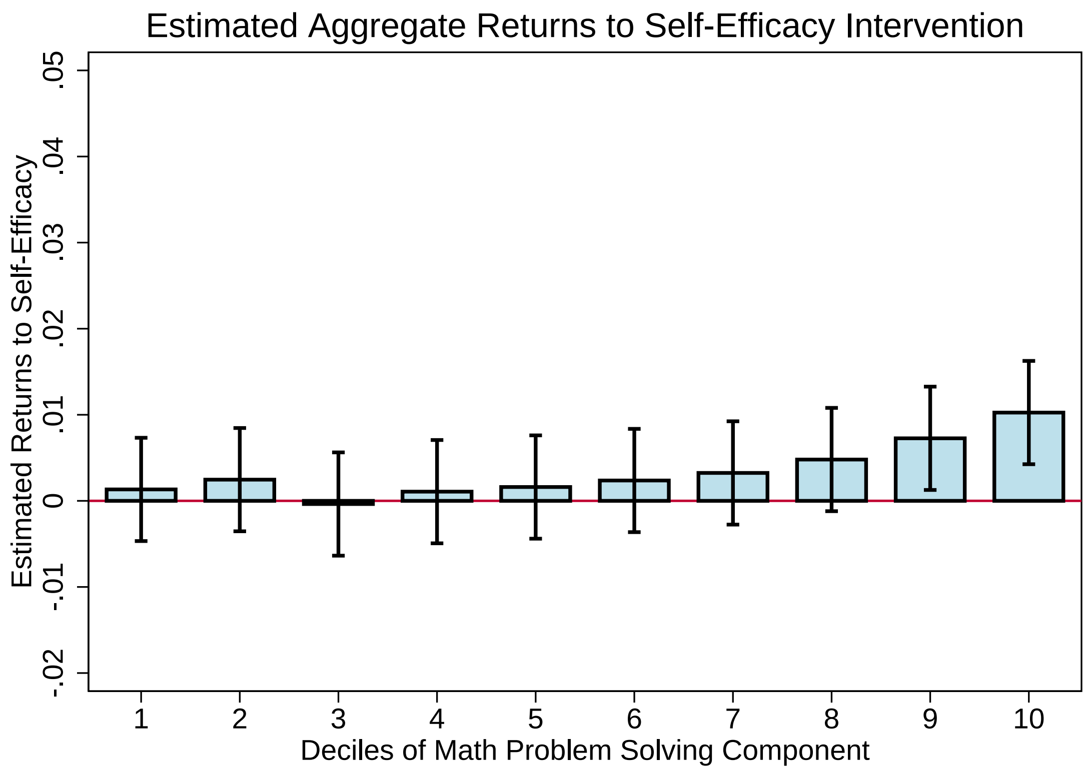

Source: Educational Longitudinal Study of 2002. Note: Figure 4.10 presents the heterogeneous impacts of the 0.25 SD self-efficacy intervention on women's STEM graduation rates across the $\theta_C$ distribution. The first panel analyzes the impact for always-takers, who are those who enroll in STEM in the baseline and continue to do so after the intervention. The panel in the top right examines the change in completion rates among compliers, who are those switching out of non-STEM enrollment into STEM enrollment. The bottom left panel analyzes the effect on those who did not enroll in STEM in the baseline or after the intervention. The bottom right panel examines the aggregate impact at each decile of the $\theta_C$ distribution by combining the effect across the three potential response groups shown in the other three panels.

**Figure 4.11:** Effect of Self-Efficacy Intervention on Wages by Response Type




Source: Educational Longitudinal Study of 2002. Note: The first panel of Figure 4.11 presents the impact of a 0.25 SD boost in $\theta_{SE}$ for the set of female college students who start in a STEM field and do so under the intervention, as well. I present heterogeneous impacts across the $\theta_C$ distribution. The second panel presents the same analysis for compliers, that is, the students who in the baseline *did not* enroll in STEM, but did so following the intervention.

**Figure 4.12:** Estimated Impacts of Self-Efficacy Intervention on Hourly Wages




Source: Educational Longitudinal Study of 2002.

Note: Figure 4.12 presents the impact of a 0.25 SD boost in $\theta_{SE}$ for all female college students included in the sample. The blue bars represent the gain in log hourly wages from the simulated policy. I present heterogeneous impacts across the $\theta_C$ distribution.

# Appendices

## A.1 Appendix Tables and Figures

**Table A.1.1:** Determinants of Enrollment Decision: Non-Cognitive Components

| | Not Enrolled | 4-Yr Inclusive | 4-Yr Selective | 4-Yr Highly Selective |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Math Test Score | -0.0817*** | -0.00384 | 0.0371*** | 0.0870*** |
| | (0.00662) | (0.00472) | (0.00617) | (0.00557) |
| English Test Score | -0.0437*** | 0.00369 | 0.0201*** | 0.0511*** |
| | (0.00644) | (0.00463) | (0.00599) | (0.00530) |
| Control Expectation | -0.0223** | 0.00590 | 0.00922 | 0.0000950 |
| | (0.00682) | (0.00475) | (0.00588) | (0.00496) |
| Instrumental Motivation | -0.0234*** | -0.00239 | 0.00811 | 0.0115** |
| | (0.00680) | (0.00452) | (0.00536) | (0.00424) |
| Action Control | -0.000620 | -0.000784 | 0.000435 | 0.0231*** |
| | (0.00783) | (0.00537) | (0.00645) | (0.00524) |
| Male | 0.120*** | -0.00170 | -0.0576*** | -0.0261*** |
| | (0.00892) | (0.00640) | (0.00808) | (0.00671) |
| White | -0.0317** | -0.0342*** | 0.0568*** | 0.0241* |
| | (0.0119) | (0.00861) | (0.0130) | (0.0113) |
| Black | -0.0528** | 0.0369*** | 0.0568** | 0.0243 |
| | (0.0167) | (0.0107) | (0.0184) | (0.0174) |
| Asian | -0.0722** | -0.0294 | 0.0438* | 0.0707*** |
| | (0.0264) | (0.0181) | (0.0222) | (0.0169) |
| Both Parents | -0.0162 | -0.00406 | 0.00954 | 0.0230** |
| | (0.0106) | (0.00759) | (0.0101) | (0.00886) |
| Family Income | -0.0185*** | 0.00469 | 0.00515 | -0.00515** |
| | (0.00356) | (0.00260) | (0.00270) | (0.00180) |
| Parents' Education | -0.0239*** | 0.000973 | 0.00994*** | 0.0212*** |
| | (0.00190) | (0.00134) | (0.00172) | (0.00147) |
| Urban | -0.0355*** | 0.0195** | 0.00414 | 0.0373*** |
| | (0.0105) | (0.00698) | (0.00906) | (0.00724) |
| Observations | | | 9,180 | |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.1.1 presents the estimated marginal effects from a multinomial logit regression, as in equation (2.1), examining the determinants of initial enrollment decisions. The omitted category is two-year college enrollment. The three different components of the non-cognitive skill index are included separately.

**Table A.1.2:** Determinants of Initial Enrollment Decisions by Gender

**Panel A.** Determinants of Enrollment Decision for Women

| | Outcome 0 | Outcome 2 | Outcome 3 | Outcome 4 |
|---|---|---|---|---|
| Math Test Score | -0.0802*** | 0.000000381 | 0.0361*** | 0.0975*** |
| | (0.00909) | (0.00667) | (0.00894) | (0.00775) |
| English Test Score | -0.0458*** | -0.00310 | 0.0294** | 0.0466*** |
| | (0.00899) | (0.00674) | (0.00903) | (0.00768) |
| Non-Cognitive Index | -0.0343*** | 0.0110* | 0.0159** | 0.0263*** |
| | (0.00598) | (0.00448) | (0.00597) | (0.00502) |
| Both Parents | -0.00459 | -0.0154 | 0.0276 | 0.0195 |
| | (0.0138) | (0.0101) | (0.0145) | (0.0120) |
| Family Income | -0.0149** | 0.00394 | 0.00459 | -0.00785** |
| | (0.00496) | (0.00371) | (0.00390) | (0.00245) |
| Parents' Education | -0.0209*** | 0.000684 | 0.0106*** | 0.0206*** |
| | (0.00246) | (0.00182) | (0.00242) | (0.00200) |
| Observations | 4,890 | | | |

**Panel B.** Determinants of Enrollment Decision for Men

| | Outcome 0 | Outcome 1 | Outcome 2 | Outcome 3 |
|---|---|---|---|---|
| Math Test Score | -0.0859*** | -0.00441 | 0.0386*** | 0.0713*** |
| | (0.00960) | (0.00656) | (0.00823) | (0.00778) |
| English Test Score | -0.0411*** | 0.0106 | 0.0109 | 0.0547*** |
| | (0.00934) | (0.00635) | (0.00768) | (0.00721) |
| Non-Cognitive Index | -0.0467*** | -0.00789 | 0.0148** | 0.0382*** |
| | (0.00675) | (0.00462) | (0.00557) | (0.00505) |
| Both Parents | -0.0322* | 0.0101 | -0.0113 | 0.0274* |
| | (0.0164) | (0.0115) | (0.0138) | (0.0131) |
| Family Income | -0.0215*** | 0.00532 | 0.00499 | -0.00154 |
| | (0.00516) | (0.00363) | (0.00370) | (0.00266) |
| Parents' Education | -0.0271*** | 0.00169 | 0.00910*** | 0.0216*** |
| | (0.00293) | (0.00201) | (0.00243) | (0.00217) |
| Observations | 4,290 | | | |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.1.2 presents the estimated marginal effects from a multinomial logit regression, as in equation (2.1), examining the determinants of initial enrollment decisions by gender. The omitted category is two-year college enrollment. The coefficients on urban status and racial categories are included in the regression, but not presented here for expositional simplicity.

**Table A.1.3:** Determinants of Academic Undermatch: Non-Cognitive Components

| | Top Test Score Decile | Top Test Score Quintile | Top Two Quintiles |
|---|---|---|---|
| | (1) | (2) | (3) |
| Math Test Score | -0.129* | -0.199*** | -0.164*** |
| | (0.0531) | (0.0307) | (0.0172) |
| English Test Score | -0.132** | -0.146*** | -0.103*** |
| | (0.0452) | (0.0257) | (0.0146) |
| Control Expectation | -0.0233 | 0.00819 | 0.00196 |
| | (0.0236) | (0.0161) | (0.0110) |
| Instrumental Motivation | -0.0149 | -0.0226 | -0.0176 |
| | (0.0177) | (0.0132) | (0.00952) |
| Action Control | -0.0374 | -0.0612*** | -0.0430*** |
| | (0.0217) | (0.0162) | (0.0117) |
| Male | 0.0638 | 0.0511* | 0.140*** |
| | (0.0329) | (0.0226) | (0.0156) |
| White | 0.00336 | -0.0136 | -0.0926*** |
| | (0.0604) | (0.0416) | (0.0253) |
| Black | 0.204 | 0.157 | -0.0553 |
| | (0.167) | (0.0878) | (0.0483) |
| Asian | -0.142 | -0.0992 | -0.101* |
| | (0.0820) | (0.0592) | (0.0410) |
| Both Parents | -0.129** | -0.0497 | -0.0253 |
| | (0.0432) | (0.0302) | (0.0203) |
| Family Income | 0.0134 | 0.00865 | -0.0000807 |
| | (0.00750) | (0.00575) | (0.00440) |
| Parents' Education | -0.0532*** | -0.0446*** | -0.0393*** |
| | (0.00753) | (0.00505) | (0.00341) |
| Urban | -0.0919** | -0.0915*** | -0.0551** |
| | (0.0348) | (0.0250) | (0.0177) |
| Constant | 1.689*** | 1.669*** | 1.318*** |
| | (0.181) | (0.115) | (0.0748) |
| Observations | 920 | 1,840 | 3,670 |
| $R^2$ | 0.143 | 0.149 | 0.145 |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.1.3 presents the estimated results from equation (2.2). Equation (2.2) estimates the determinants of academic undermatch for students at different points of the math test score distribution.

**Table A.1.4:** Determinants of Academic Undermatch: Excluding Non-Enrolled Students

| | Top Test Score Decile | Top Test Score Quintile | Top Two Quintiles |
|---|---|---|---|
| | (1) | (2) | (3) |
| Math Test Score | -0.0940 | -0.158*** | -0.134*** |
| | (0.0554) | (0.0328) | (0.0172) |
| English Test Score | -0.128** | -0.134*** | -0.0712*** |
| | (0.0465) | (0.0274) | (0.0150) |
| Non-Cognitive Index | -0.0636*** | -0.0775*** | -0.0469*** |
| | (0.0190) | (0.0132) | (0.00849) |
| Male | 0.0125 | 0.0368 | 0.0932*** |
| | (0.0333) | (0.0238) | (0.0158) |
| White | 0.0529 | 0.0147 | -0.0933*** |
| | (0.0654) | (0.0464) | (0.0267) |
| Black | 0.316 | 0.203* | -0.0553 |
| | (0.179) | (0.0990) | (0.0501) |
| Asian | -0.137 | -0.138* | -0.129** |
| | (0.0874) | (0.0658) | (0.0424) |
| Both Parents | -0.134** | -0.0616 | -0.0184 |
| | (0.0449) | (0.0327) | (0.0210) |
| Family Income | 0.0153* | 0.0133* | 0.00578 |
| | (0.00761) | (0.00619) | (0.00446) |
| Parents' Education | -0.0497*** | -0.0429*** | -0.0313*** |
| | (0.00784) | (0.00548) | (0.00348) |
| Urban | -0.0981** | -0.0999*** | -0.0599*** |
| | (0.0358) | (0.0268) | (0.0179) |
| Constant | 1.499*** | 1.488*** | 0.994*** |
| | (0.188) | (0.126) | (0.0764) |
| Observations | 840 | 1,630 | 3,120 |
| $R^2$ | 0.133 | 0.133 | 0.106 |

Source: Educational Longitudinal Study of 2002. Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.1.4 presents the estimated results from equation (2.2). Equation (2.2) estimates the determinants of academic undermatch for students at different points of the math test score distribution. This table excludes students who had not enrolled in a higher education institution by age 20. In the first two columns, students are classified to undermatch if they are not enrolled in a highly-selective college. In the third column, undermatched students are those who have not enrolled in a selective or highly-selective institution by age 20.

## B.1 Appendix Tables and Figures

**Table B.1.1:** Non-Linear Returns to Mathematical Ability

| | (1) | (2) |
|---|---|---|
| $P_{51} - P_{80}$ | 0.210*** | 0.182*** |
| | (0.0007) | (0.0008) |
| $P_{81} - P_{90}$ | 0.415*** | 0.362*** |
| | (0.0009) | (0.0009) |
| $P_{91} - P_{95}$ | 0.583*** | 0.519*** |
| | (0.0009) | (0.0009) |
| $P_{96} - P_{99}$ | 0.770*** | 0.697*** |
| | (0.0009) | (0.0009) |
| $P_{100}$ | 1.011*** | 0.930*** |
| | (0.0009) | (0.0009) |
| Language | | X |
| Year FE | X | X |
| $R^2$ | 0.125 | 0.126 |
| Observations | 10,170,432 | |
| Individual Observations | 243,267 | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples. SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized. Wages are measured monthly in 2010 Real CLP in the highest paid job.

**Table B.1.2:** Non-Linear Returns to Mathematical Ability (Monthly Wages)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $P_{51} - P_{80}$ | 0.185*** | 0.129*** | 0.129*** | 0.130*** | 0.125*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| $P_{81} - P_{90}$ | 0.389*** | 0.281*** | 0.277*** | 0.277*** | 0.253*** |
| | (0.0009) | (0.0009) | (0.0009) | (0.0009) | (0.0009) |
| $P_{91} - P_{95}$ | 0.569*** | 0.421*** | 0.414*** | 0.405*** | 0.367*** |
| | (0.0011) | (0.0011) | (0.0011) | (0.0011) | (0.0011) |
| $P_{96} - P_{99}$ | 0.765*** | 0.584*** | 0.575*** | 0.540*** | 0.495*** |
| | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) |
| $P_{100}$ | 1.010*** | 0.787*** | 0.777*** | 0.695*** | 0.644*** |
| | (0.0025) | (0.0025) | (0.0025) | (0.0025) | (0.0025) |
| Year FE | X | X | X | X | X |
| Years of Ed. | | X | | | |
| Degrees Received | | | X | | |
| University Quality | | | | X | |
| Field of Degree | | | | | X |
| $R^2$ | 0.165 | 0.233 | 0.234 | 0.236 | 0.244 |
| Observations | | | 10,170,432 | | |
| Individual Observations | | | 243,267 | | |

Note: SE clustered at the individual level. * p<0.05, ** p<0.01, *** p<0.001. SIMCE scores from 2001 and 2003 $10^{th}$ grade samples. SIES Higher Education Degrees — 2007-2016. Unemployment Insurance: 2002-2016. Ability measures are standardized. Wages are measured monthly in 2010 Real CLP in the highest paid job.

Source: Educational Longitudinal Study of 2002.

Note: Table B.1.3 displays the estimation results from the measurement system of test scores separately by gender. The dependent variable is the normalized (0,1) test score. Both Par. represents a dummy variable for whether the individual lives in a two-parent family and parents' education is a continuous variable for the surveyed parent's years of education completed. Urban is a dummy variable indicating whether the family resides in an urban area. Father in STEM and mother in STEM are dummy variables which equal one if the father and/or mother work in a STEM occupation, respectively. Standard errors are in parentheses. Various loadings are normalized to one for model identification.

**Table B.1.3:** Factor Loadings: Measurement System for Women

| | BY Math | F1 Math | SAT Math | Math GPA | BY SE | F1 SE | BY Engl | Engl GPA | SAT Read |
|---|---|---|---|---|---|---|---|---|---|
| Constant | -1.26 | -1.26 | -1.20 | -0.18 | -0.41 | -0.28 | -1.58 | -0.09 | -1.60 |
| | (0.13) | (0.12) | (0.12) | (0.12) | (0.11) | (0.10) | (0.13) | (0.12) | (0.13) |
| White | 0.01 | -0.10 | -0.19 | -0.01 | -0.02 | 0.02 | 0.28 | -0.00 | 0.25 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Black | -0.82 | -0.93 | -0.85 | -0.67 | 0.02 | -0.02 | -0.44 | -0.74 | -0.52 |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.06) | (0.06) | (0.06) | (0.05) | (0.06) |
| Both Par. | 0.12 | 0.13 | 0.11 | 0.08 | 0.09 | 0.04 | 0.11 | 0.06 | 0.01 |
| | (0.04) | (0.03) | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.03) | (0.04) |
| Fam. Income | -0.00 | -0.01 | -0.02 | -0.02 | -0.01 | -0.01 | 0.01 | -0.01 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Parents' Ed. | 0.07 | 0.08 | 0.09 | 0.03 | 0.02 | 0.01 | 0.08 | 0.03 | 0.09 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Urban | -0.12 | -0.05 | -0.08 | -0.15 | -0.00 | 0.00 | -0.09 | -0.17 | -0.06 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Father in STEM | 0.15 | 0.11 | 0.10 | -0.00 | 0.01 | 0.07 | 0.16 | 0.05 | 0.12 |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) |
| Mother in STEM | 0.18 | 0.18 | 0.26 | 0.15 | 0.23 | 0.13 | 0.25 | 0.03 | 0.25 |
| | (0.10) | (0.10) | (0.10) | (0.10) | (0.10) | (0.08) | (0.10) | (0.09) | (0.09) |
| $\theta_C$ | 0.99 | 1.10 | 1.00 | 0.69 | 0.42 | 0.34 | 0.70 | 0.20 | 0.67 |
| | (0.01) | (0.01) | (0.00) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| $\theta_{SE}$ | 0.00 | 0.00 | 0.00 | 0.29 | 0.67 | 1.00 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.02) | (0.02) | (0.00) | (0.00) | (0.00) | (0.00) |
| $\theta_R$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.51 | 0.57 | 1.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.03) | (0.02) | (0.00) |
| Precision | 4.31 | 9.60 | 5.64 | 1.88 | 1.90 | 10.76 | 2.44 | 1.67 | 10.82 |
| | (0.11) | (0.41) | (0.16) | (0.04) | (0.05) | (1.77) | (0.07) | (0.03) | (2.96) |

**Table B.1.4:** Factor Loadings: Measurement System for Men

| | BY Math | F1 Math | SAT Math | Math GPA | BY SE | F1 SE | BY Engl | Engl GPA | SAT Read |
|---|---|---|---|---|---|---|---|---|---|
| Constant | -1.16 | -0.83 | -1.11 | -0.46 | 0.07 | 0.08 | -1.47 | -0.73 | -1.91 |
| | (0.15) | (0.14) | (0.15) | (0.15) | (0.12) | (0.08) | (0.14) | (0.17) | (0.15) |
| White | -0.09 | -0.08 | -0.17 | -0.04 | -0.06 | 0.01 | 0.13 | -0.06 | 0.07 |
| | (0.03) | (0.03) | (0.04) | (0.04) | (0.03) | (0.02) | (0.04) | (0.04) | (0.03) |
| Black | -0.94 | -0.96 | -0.97 | -0.72 | -0.11 | -0.07 | -0.61 | -0.74 | -0.66 |
| | (0.07) | (0.07) | (0.06) | (0.07) | (0.06) | (0.04) | (0.07) | (0.07) | (0.06) |
| Both Par. | 0.12 | 0.06 | 0.06 | 0.06 | 0.01 | 0.01 | 0.11 | 0.02 | 0.06 |
| | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.02) | (0.05) | (0.05) | (0.04) |
| Fam. Income | -0.01 | -0.01 | -0.02 | -0.00 | -0.00 | -0.00 | -0.02 | 0.01 | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.01) | (0.01) | (0.01) |
| Parents' Ed. | 0.09 | 0.08 | 0.10 | 0.03 | 0.01 | 0.01 | 0.09 | 0.03 | 0.13 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.01) | (0.01) | (0.01) |
| Urban | 0.01 | 0.02 | 0.05 | -0.12 | 0.05 | -0.01 | 0.06 | -0.15 | 0.04 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.04) | (0.03) |
| Father in STEM | 0.20 | 0.17 | 0.06 | 0.14 | 0.06 | 0.00 | 0.16 | 0.12 | 0.16 |
| | (0.06) | (0.06) | (0.06) | (0.06) | (0.05) | (0.03) | (0.06) | (0.06) | (0.05) |
| Mother in STEM | 0.12 | 0.08 | 0.25 | -0.15 | 0.12 | 0.03 | 0.27 | 0.10 | 0.26 |
| | (0.10) | (0.11) | (0.11) | (0.13) | (0.09) | (0.06) | (0.11) | (0.12) | (0.12) |
| $\theta_C$ | 1.01 | 1.06 | 1.00 | 0.59 | 0.31 | 0.14 | 0.69 | 0.25 | 0.67 |
| | (0.01) | (0.01) | (0.00) | (0.02) | (0.02) | (0.01) | (0.02) | (0.02) | (0.02) |
| $\theta_{SE}$ | 0.00 | 0.00 | 0.00 | 0.25 | 0.59 | 1.00 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) |
| $\theta_R$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.56 | 0.54 | 1.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.06) | (0.03) | (0.00) |
| Precision | 4.52 | 9.99 | 3.63 | 1.61 | 2.08 | 44.97 | 2.03 | 1.27 | 6.49 |
| | (0.14) | (0.51) | (0.10) | (0.04) | (0.05) | (5.35) | (0.08) | (0.03) | (1.55) |

Source: Educational Longitudinal Study of 2002.
Note: Table B.1.4 displays the estimation results from the measurement system of test scores separately by gender. The dependent variable is the normalized (0,1) test score. Both Par. represents a dummy variable for whether the individual lives in a two-parent family and parents' education is a continuous variable for the surveyed parent's years of education completed. Urban is a dummy variable indicating whether the family resides in an urban area. Father in STEM and mother in STEM are dummy variables which equal one if the father and/or mother work in a STEM occupation, respectively. Standard errors are in parentheses. Various loadings are normalized to one for model identification.

**Table B.1.5:** Participation in Math-Intensive STEM Majors

| | Initial Major | | Final Major | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| | (1) | (2) | (3) | (4) |
| White | -0.009 | 0.021 | -0.014 | -0.038 |
| | (0.015) | (0.031) | (0.016) | (0.033) |
| Parental Education | 0.003 | -0.004 | 0.000 | -0.005 |
| | (0.002) | (0.006) | (0.003) | (0.006) |
| Log Family Income | 0.006** | 0.003 | 0.005* | 0.004 |
| | (0.003) | (0.006) | (0.003) | (0.007) |
| Father in Field | 0.020 | 0.074 | 0.024 | 0.074 |
| | (0.027) | (0.046) | (0.028) | (0.049) |
| Mother in Field | 0.021 | 0.073 | 0.019 | -0.028 |
| | (0.043) | (0.101) | (0.044) | (0.099) |
| Low Math | | | | |
| $x$ Medium-SE | -0.006 | 0.094** | 0.019 | 0.126*** |
| | (0.009) | (0.041) | (0.014) | (0.042) |
| $x$ High-SE | 0.039 | 0.118** | 0.065** | 0.172*** |
| | (0.025) | (0.054) | (0.030) | (0.059) |
| Medium Math | | | | |
| $x$ Low-SE | 0.009 | 0.034 | 0.012 | 0.059 |
| | (0.016) | (0.037) | (0.012) | (0.038) |
| $x$ Medium-SE | 0.016 | 0.086** | 0.022 | 0.156*** |
| | (0.015) | (0.041) | (0.016) | (0.044) |
| $x$ High-SE | 0.070*** | 0.264*** | 0.089*** | 0.305*** |
| | (0.025) | (0.053) | (0.027) | (0.054) |
| High Math | | | | |
| $x$ Low-SE | -0.004 | 0.021 | 0.027 | 0.074 |
| | (0.022) | (0.059) | (0.026) | (0.063) |
| $x$ Medium-SE | 0.062** | 0.155*** | 0.057** | 0.268*** |
| | (0.024) | (0.045) | (0.022) | (0.048) |
| $x$ High-SE | 0.101*** | 0.396*** | 0.126*** | 0.464*** |
| | (0.022) | (0.046) | (0.024) | (0.046) |
| $R^2$ | 0.039 | 0.096 | 0.044 | 0.113 |
| Observations | 1,340 | 1,090 | 1,340 | 1,090 |

Source: Educational Longitudinal Study of 2002. Note: * p<0.10, ** p<0.05, *** p<0.01. Note: This sample includes students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey, had completed a college degree by 2012, yet not a graduate degree and had reported a positive hourly wage in 2012. Students are required to have reported grades/scores for all the baseline math test score and self-efficacy measure, individual and family background characteristics. The estimated regression is a linear probability model with STEM enrollment and STEM graduation as the outcome variables, respectively. This regression is separately estimated by gender. For each gender, I divide the sample in tertiles of the math test score and of the self-efficacy measure and then interact these categories creating nine separate gender-specific "math skill bins." The omitted category represents individuals in the bottom math test score and math self-efficacy tertiles.

**Table B.1.6:** Productivity of Latent Ability Across Initial Majors by Gender

| | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|
| Initial Major | STEM Enrollee | | Other Majors | | STEM Enrollee | | Other Majors | |
| Outcome | STEM | College | STEM | College | STEM | College | STEM | Graduate |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\theta_C$ | 0.136*** | 0.190*** | 0.011*** | 0.076*** | 0.154*** | 0.025*** | 0.042*** | 0.066*** |
| | (0.012) | (0.010) | (0.001) | (0.002) | (0.006) | (0.005) | (0.001) | (0.002) |
| $\theta_{SE}$ | 0.0989*** | 0.184*** | 0.005*** | -0.018*** | -0.009 | 0.029*** | 0.023*** | 0.012*** |
| | (0.010) | (0.008) | (0.001) | (0.002) | (0.006) | (0.005) | (0.001) | (0.002) |
| $\theta_R$ | 0.0528*** | -0.113*** | -0.002*** | 0.031*** | 0.003 | 0.053*** | -0.039*** | -0.022*** |
| | (0.0117) | (0.009) | (0.001) | (0.002) | (0.009) | (0.007) | (0.001) | (0.004) |

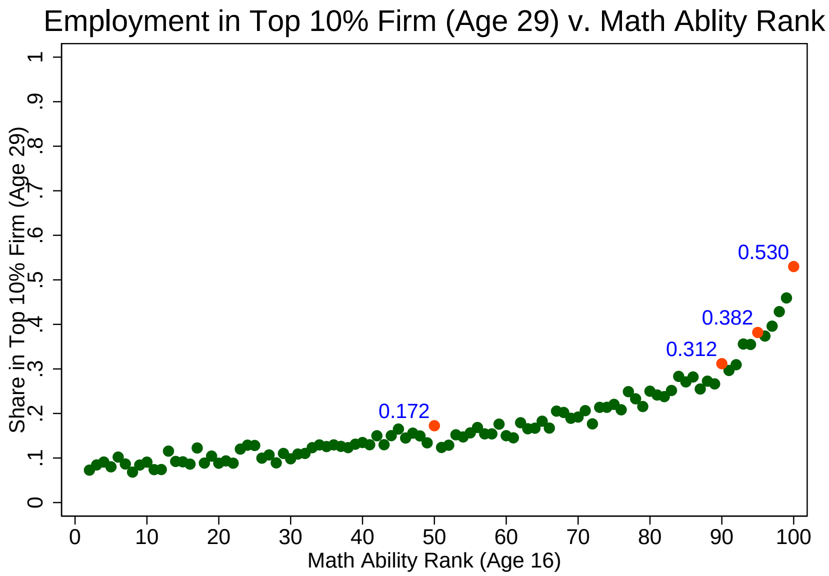
Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: Table B.1.6 estimates the effect of an increase in each of the three latent factor dimensions on the probability of STEM completion (odd columns) and college completion (even columns) for men and women, depending on their initial major choice (columns (1)-(2) and (5)-(6) explore these patterns for STEM enrollees, whereas (3)-(4) and (7)-(8) do so for non-enrollees). The results are estimated following a linear probability model using simulated model parameters.

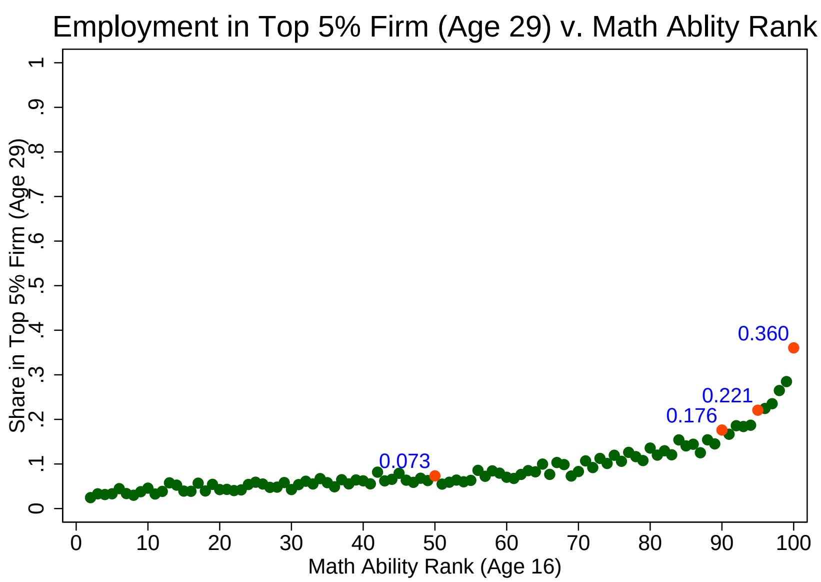
**Table B.1.7:** Sorting into College Continuation by Initial Major and Gender

| STEM Enrollment Rates | | | | |
|---|---|---|---|---|
| Baseline<br>(1) | Problem Solving Comp.<br>(2) | Self-Efficacy Comp.<br>(3) | Joint Compensation<br>(4) | Male Share<br>(5) |
| 0.045 | 0.053 | 0.053 | 0.062 | 0.179 |
| (0.001) | (0.001)*** | (0.001)*** | (0.001)*** | |
| **STEM Graduation Rates** | | | | |
| Baseline<br>(1) | Problem Solving Comp.<br>(2) | Self-Efficacy Comp.<br>(3) | Joint Compensation<br>(4) | Male Share<br>(5) |
| 0.029 | 0.038 | 0.037 | 0.047 | 0.158 |
| (0.001) | (0.001)*** | (0.001)*** | (0.001)*** | |

Source: Educational Longitudinal Study of 2002. Note: * p<0.10, ** p<0.05, *** p<0.01. Note: Table B.1.7 explores changing female STEM participation under different counterfactual compensation policies described in the paper. Stars on columns (2)-(4) represent the significance of the t-tes examining whether participation rates have changed under the counterfactual scenarios.

**Figure 13:** Alternative Firm Quality Measures



## C.1 Appendix Tables

**Table C.1.1:** Sample Restrictions

| Sample Restriction | Number of Observations |
|---|---|
| Full Sample | 16,200 |
| Baseline Respondents | 15,890 |
| Baseline Test Takers | 13,440 |
| SAT and High School Grades | 12,390 |
| HS Graduates by Second Follow-Up | 12,510 |
| Enrolled in Four-Year College | 4,630 |
| Missing Initial Majors | 4,600 |
| Missing Final Attainment | 4,520 |

Source: Educational Longitudinal Study of 2002.
Note: Table C.1.1 shows the sample restrictions imposed on ELS data. I require respondents to have participated in the baseline survey and to have taken the two baseline math and reading scores, the follow-up math test and report an ACT/SAT score. The sample is comprised students who had graduated high school by 2004 (age 18/19) and were enrolled in four-year college by age 20/21. I drop students who do not report information on their college major at enrollment as well as those not participating in the final follow-up survey.

**Table C.1.2:** Summary Statistics by Major

**Panel A.** STEM Enrollment Patterns by Gender

| | Women | | | Men | | |
|---|---|---|---|---|---|---|
| | STEM (1) | Non-STEM (2) | Difference (3) | STEM (4) | Non-STEM (5) | Difference (6) |
| Share in Major | 0.045 | 0.955 | | 0.179 | 0.821 | |
| Baseline Math Test | 0.274 | -0.103 | 0.377*** | 0.539 | 0.095 | 0.444*** |
| | (1.033) | (0.942) | (0.091) | (0.963) | (0.976) | (0.057) |
| Baseline Self-Efficacy | 0.331 | -0.157 | 0.488*** | 0.607 | 0.122 | 0.485*** |
| | (0.917) | (0.994) | (0.095) | (0.869) | (0.953) | (0.055) |

**Panel B.** Final Outcomes for Women by Initial Degree

| | STEM Enrollees | | | Non-STEM Enrollees | | |
|---|---|---|---|---|---|---|
| | STEM (1) | Other (2) | Dropout (3) | STEM (4) | Other (5) | Dropout (6) |
| Final Outcome | 0.442 | 0.327 | 0.230 | 0.013 | 0.784 | 0.202 |
| Baseline Math Test | 0.666 | 0.095*** | -0.225*** | 0.434 | -0.026*** | -0.435*** |
| | | (0.209) | (0.212) | | (0.163) | (0.177) |
| Baseline Self-Efficacy | 0.637 | 0.309*** | -0.224*** | 0.070 | -0.133*** | -0.263*** |
| | | (0.184) | (0.198) | | (0.176) | (0.180) |

**Panel C.** Final Outcomes for Men by Initial Degree

| | STEM Enrollees | | | Non-STEM Enrollees | | |
|---|---|---|---|---|---|---|
| | STEM (1) | Other (2) | Dropout (3) | STEM (4) | Other (5) | Dropout (6) |
| Final Outcome | 0.614 | 0.211 | 0.175 | 0.058 | 0.784 | 0.218 |
| Baseline Math Test | 0.730 | 0.282*** | 0.183*** | 0.555 | 0.125*** | -0.128*** |
| | | (0.120) | (0.130) | | (0.100) | (0.116) |
| Baseline Self-Efficacy | 0.698 | 0.522 | 0.388* | 0.514 | 0.104** | 0.077*** |
| | | (0.113) | (0.121) | | (0.101) | (0.108) |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. Majors are categorized by STEM and non-STEM fields. All test score measures are normalized (0,1) for comparability. Table C.1.2 shows sorting patterns across initial and final majors on the baseline math test score and self-efficacy. In Panels B and C, stars on columns (2)-(3) and (5)-(6) represent the significance of the t-test examining whether baseline test scores are differences among switchers and completers and dropouts and completers. To interpret the tests, note that in those columns, I report the baseline value of each corresponding test. The difference is given by the subtraction off the column (1) and column (4) values, respectively The values in parentheses in Panels B and C represent the standard errors of these tests.

**Table C.1.3:** Gender Gaps in in Math-Intensive STEM Major Participation

| | Initial Choice | | | Final Choice | | |
|---|---|---|---|---|---|---|
| | Baseline | Linear | Non-Linear | Baseline | Linear | Non-Linear |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gender Wage Gap | 0.161*** | 0.135*** | 0.134*** | 0.195*** | 0.163*** | 0.162*** |
| | (0.013) | (0.014) | (0.014) | (0.013) | (0.014) | (0.014) |
| Math Test Score | | 0.012*** | | | 0.016*** | |
| | | (0.003) | | | (0.003) | |
| | | [0.075] | | | [0.082] | |
| Math Self-Efficacy | | 0.012*** | | | 0.014*** | |
| | | (0.003) | | | (0.003) | |
| | | [0.075] | | | [0.072] | |
| Reading Test Score | | 0.002* | 0.002* | | 0.002* | 0.002* |
| | | (0.001) | (0.001) | | (0.001) | (0.001) |
| | | [0.012] | [0.012] | | [0.011] | [0.011] |
| Test Score × SE Bins (25) | | | 0.026*** | | | 0.031*** |
| | | | (0.005) | | | (0.005) |
| | | | [0.161] | | | [0.164] |

Source: Educational Longitudinal Study of 2002.
Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey and who had not completed a graduate degree in the final round. I examine gender gaps in STEM enrollment and completion. Table C.1.3 examines the contribution of baseline test scores to gender gaps in STEM participation, both at enrollment and graduation. All test score measures are normalized (0,1) for comparability. All regressions include students' race, family composition, parental income, parents' education, region of residence dummy variables and urban residence as control variables. Bracketed terms indicate the share of the gender gap in majors which is explained by each component of observed test scores. The "Test Score x SE Bins (25)" row denotes a semi-parametric model, in which I placed all students in one of five quintiles of baseline math test scores and self-efficacy. After interacting these categories, students were placed in one of 25 bins.

**Table C.1.4:** Hourly Wages by Major

| | Initial Major | | | Final Major | | |
|---|---|---|---|---|---|---|
| | Women | Men | Difference | Women | Men | Difference |
| Hourly Wages | (1) | (2) | (3) | (4) | (5) | (6) |
| Math-Intensive STEM | 2.843 | 3.034 | -0.191*** | 2.965 | 3.147 | -0.182*** |
| | (0.504) | (0.486) | (0.067) | (0.517) | (0.489) | (0.069) |
| Life Sciences | 2.673 | 2.757 | -0.085 | 2.705 | 2.691 | 0.014 |
| | (0.480) | (0.482) | (0.067) | (0.447) | (0.489) | (0.055) |
| Business | 2.840 | 2.986 | -0.145*** | 2.893 | 3.047 | -0.154*** |
| | (0.433) | (0.513) | (0.046) | (0.447) | (0.489) | (0.041) |
| Health | 2.951 | 2.810 | 0.141 | 3.149 | 3.045 | 0.104 |
| | (0.453) | (0.472) | (0.087) | (0.376) | (0.489) | (0.084) |
| Other | 2.715 | 2.732 | -0.016 | 2.749 | 2.748 | -0.001 |
| | (0.457) | (0.549) | (0.028) | (0.472) | (0.489) | (0.032) |
| Not Declared | 2.737 | 2.778 | -0.041 | | | |
| | (0.512) | (0.515) | (0.044) | | | |
| Not Graduated | | | | 2.626 | 2.665 | -0.039 |
| | | | | (0.517) | (0.489) | (0.034) |

Source: Educational Longitudinal Study of 2002.
Note: * p<0.10, ** p<0.05, *** p<0.01.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. Students are required to have reported grades/scores for all the test scores presented above. Table C.1.4 presents average hourly wages' by students' initial majors and final outcomes, by gender. The first three panels include all college enrollees who had not completed a graduate degree by 2012. The final three columns report wages for college graduates in each node (including non-completers in the last row) who had not completed a graduate degree in 2012. Wages are reported as natural logarithms.

**Table C.1.5:** Variables Used in Implementation of the Model

| | Measurement System | Initial Major | Dropout | Final Major | Graduate School | Employment | Wage Eq. |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Constant | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Race | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Family Composition | Yes | Yes | Yes | Yes | | | |
| Parents' Education | Yes | Yes | Yes | Yes | | | |
| Parents' Occupation | Yes | Yes | Yes | Yes | | | |
| Family Income | Yes | Yes | Yes | Yes | | | |
| Region of Residence (BY) | Yes | Yes | | | | | |
| Local Share of Graduates ($m_1$ | | Yes | | | | | |
| College Unemployment Rate (F2) | | | Yes | | | | |
| Unemployment by Major (F2) | | | | Yes | | | |
| Share in Graduate School (F3) | | | | | Yes | | |
| Unemployment by Major (F3) | | | | | Yes | | |
| Region of Residence (F3) | | | | | | Yes | Yes |
| Latent Ability | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Source: Educational Longitudinal Study of 2002.
Note: I show the variables used in the empirical model. In the measurement system, I use three math test scores, students' math GPA in high school, two measures of mathematical self-efficacy, two reading test scores and students English GPA in high school. Initial major choices are defined in six categories: math-intensive STEM, life sciences, business, health-related fields, other majors and not declaring a major. The dropout decision encompasses continuing in four-year college or not. The final major decision involves staying in the original major or switching to a different field. The graduate school decision involves completing a graduate degree by age 26. Hourly wages are measured in logs at age 26.

**Table C.1.6:** Choice Equations: Loadings for Women. STEM Decisions.

| | Initial Major | Continuation | Major Switch | Master's | Work Decision | Wages |
|---|---|---|---|---|---|---|
| Constant | -2.28 | 3.71 | -0.47 | -4.79 | 8.36 | 2.84 |
| | (0.42) | (1.65) | (0.82) | (2.46) | (2.79) | (0.13) |
| White | -0.08 | 0.53 | 0.13 | 0.02 | -0.08 | 0.20 |
| | (0.12) | (0.26) | (0.21) | (0.44) | (0.68) | (0.14) |
| Black | -0.04 | | | | | |
| | (0.19) | | | | | |
| Both Parents | -0.14 | -0.62 | 0.50 | 0.34 | | |
| | (0.11) | (0.33) | (0.24) | (0.60) | | |
| Fam. Income | 0.07 | 0.03 | 0.10 | 0.11 | | |
| | (0.03) | (0.06) | (0.05) | (0.15) | | |
| Parents' Ed. | 0.04 | -0.10 | | | | |
| | (0.02) | (0.05) | | | | |
| Urban | 0.29 | | | | | |
| | (0.09) | | | | | |
| $\theta_C$ | 0.42 | 0.83 | 0.30 | 0.70 | 0.40 | 0.09 |
| | (0.07) | (0.19) | (0.18) | (0.37) | (0.50) | (0.07) |
| $\theta_{SE}$ | 0.43 | 0.59 | 0.39 | 0.11 | 0.41 | -0.19 |
| | (0.11) | (0.26) | (0.20) | (0.37) | (0.15) | (0.13) |
| $\theta_R$ | -0.14 | -0.55 | 0.20 | -0.68 | -0.94 | 0.00 |
| | (0.14) | (0.34) | (0.31) | (0.65) | (1.07) | (0.00) |
| Local Share in Major | 0.02 | | | | | |
| | (0.05) | | | | | |
| Local UN (College) | | -0.116 | | | | |
| | | (0.082) | | | | |
| Local UN Rate in Major | | | -0.050 | | -0.023 | |
| | | | (0.027) | | (0.004) | |
| Local Share in Master's | | | | 0.161 | | |
| | | | | (0.149) | | |
| Precision | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.60 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.54) |

Source: Educational Longitudinal Study of 2002.
Note: Table C.1.6 displays the estimation results from choice and wage equations for women. The second column shows the estimated parameters associated with an initial major in a math-intensive STEM field. The local share in major variable represents the share of students enrolled in student $i$'s local four-year college(s) who completed a math-intensive STEM major. The third column is the continuation decision for women initially enrolled in STEM and 'Local UN (College)' is the average unemployment share in student $i$'s commuting zone of residence in the first follow-up survey. The fourth column represents the decision to stay in a STEM field or to switch to a different major for these students and the 'Local UN Rate in Major' captures local the unemployment rate for female STEM college graduates in student $i$'s commuting zone in the first follow-up survey. The Master's decision column considers the graduate degree decision and the 'Local Share in Master's' variable represents the share of college graduates aged 25-34 who have also obtained a graduate degree in person $i$'s commuting zone of residence in the final survey round. The Work Decision column represents the employment decision and the 'Local UN Rate in Major' variable captures local the unemployment rate for female STEM college graduates in student $i$'s commuting zone in final survey round. Both Par. represents a dummy variable for whether the individual lives in a two-parent family and parents' education is a continuous variable for the surveyed parent's years of education completed. Urban is a dummy variable indicating whether the family resides in an urban area. The male coefficients are similar, as are those for other majors. These are not presented for presentation simplicity and available upon request.

**Table C.1.7:** Goodness of Fit: Educational Choices

*__Panel A.__ Initial Enrollment for Men and Women

| | Women | | Men | |
|---|---|---|---|---|
| | Observed (1) | Simulated (2) | Observed (3) | Simulated (4) |
| STEM | 0.045 | 0.044 | 0.179 | 0.177 |
| Life Sciences | 0.077 | 0.078 | 0.071 | 0.071 |
| Business | 0.116 | 0.116 | 0.181 | 0.184 |
| Health | 0.135 | 0.136 | 0.036 | 0.038 |
| Other | 0.470 | 0.468 | 0.340 | 0.337 |
| Not Declared | 0.158 | 0.158 | 0.193 | 0.194 |
| Goodness of Fit (p-value) | 0.781 | | 0.823 | |

*__Panel B.__ Final Majors among STEM Enrollees

| | Women | | Men | |
|---|---|---|---|---|
| | Observed (1) | Simulated (2) | Observed (3) | Simulated (4) |
| Graduate | 0.442 | 0.429 | 0.614 | 0.607 |
| Switch | 0.327 | 0.339 | 0.211 | 0.214 |
| Dropout | 0.231 | 0.232 | 0.175 | 0.179 |
| Goodness of Fit (p-value) | 0.525 | | 0.612 | |

*__Panel C.__ Final Majors among Non-Declared Students

| | Women | | Men | |
|---|---|---|---|---|
| | Observed (1) | Simulated (2) | Observed (3) | Simulated (4) |
| STEM | 0.035 | 0.036 | 0.093 | 0.096 |
| Life Sciences | 0.093 | 0.098 | 0.070 | 0.067 |
| Business | 0.151 | 0.151 | 0.186 | 0.181 |
| Health | 0.078 | 0.078 | 0.021 | 0.025 |
| Other | 0.418 | 0.410 | 0.344 | 0.343 |
| Not Declared | 0.224 | 0.227 | 0.287 | 0.289 |
| Goodness of Fit (p-value) | 0.745 | | 0.794 | |

Source: Educational Longitudinal Study of 2002.

Note: Table C.1.7 examines goodness of fit of educational decisions in the discrete choice model. Goodness of fit is tested using a $\chi^2$ test where the Null Hypothesis is $Model = Data$. Observed majors at enrollment and graduation follow from the sample described in Section 4.2. Simulated results come from the 200,000 observations simulated for each gender using estimated model parameters.

**Table C.1.8:** Goodness of Fit: Labor Market Outcomes

*__Panel A.__ Initial Major Choices

| | Employment | | | Hourly Wages | | |
|---|---|---|---|---|---|---|
| | Observed | Simulated | Difference | Observed | Simulated | Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| STEM | 0.921 | 0.901 | -0.020 | 2.843 | 2.814 | 0.029* |
| Life Sciences | 0.927 | 0.908 | -0.019 | 2.673 | 2.686 | 0.013 |
| Business | 0.955 | 0.951 | -0.004 | 2.841 | 2.836 | -0.005 |
| Health | 0.967 | 0.962 | -0.005 | 2.951 | 2.943 | -0.008 |
| Other | 0.942 | 0.938 | -0.004 | 2.715 | 2.723 | 0.008 |
| Not Declared | 0.947 | 0.937 | -0.010 | 2.737 | 2.749 | 0.012 |

*__Panel B.__ Final Majors among STEM Enrollees

| | Employment | | | Hourly Wages | | |
|---|---|---|---|---|---|---|
| | Observed | Simulated | Difference | Observed | Simulated | Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Graduate | 0.884 | 0.852 | -0.032* | 2.970 | 2.952 | 0.018 |
| Switch | 0.900 | 876 | -0.024 | 2.772 | 2.790 | 0.018 |
| Dropout | 1.000 | 1.000 | 0.000 | 2.690 | 2.708 | 0.018 |

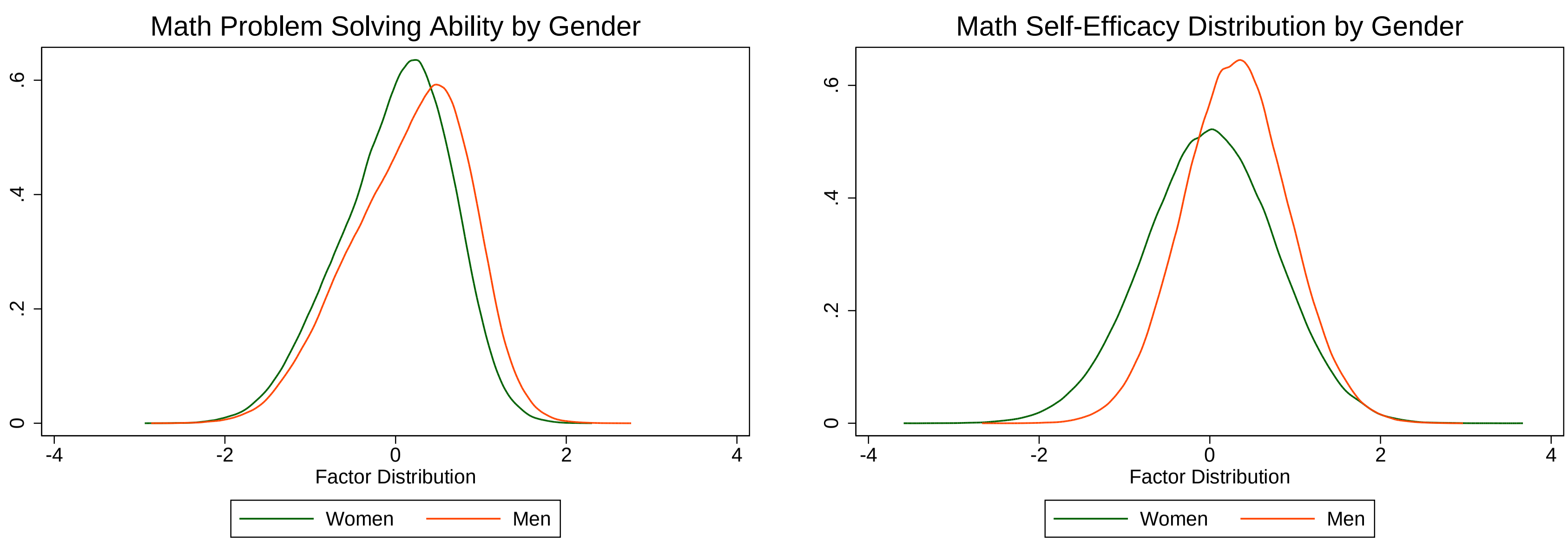
*__Panel C.__ Final Majors among Non-Declared Students

| | Employment | | | Hourly Wages | | |
|---|---|---|---|---|---|---|
| | Observed | Simulated | Difference | Observed | Simulated | Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| STEM | 0.909 | 0.901 | -0.008 | 2.892 | 2.868 | -0.026 |
| Life Sciences | 0.911 | 0.883 | -0.028 | 2.720 | 2.724 | 0.004 |
| Business | 0.944 | 0.931 | -0.013 | 2.886 | 2.890 | 0.004 |
| Health | 1.000 | 0.997 | -0.003 | 2.986 | 3.060 | 0.072** |
| Other | 0.947 | 0.945 | -0.002 | 2.756 | 2.758 | 0.002 |
| Not Declared | 0.955 | 0.947 | -0.008 | 2.562 | 2.534 | -0.028 |

Source: Educational Longitudinal Study of 2002.

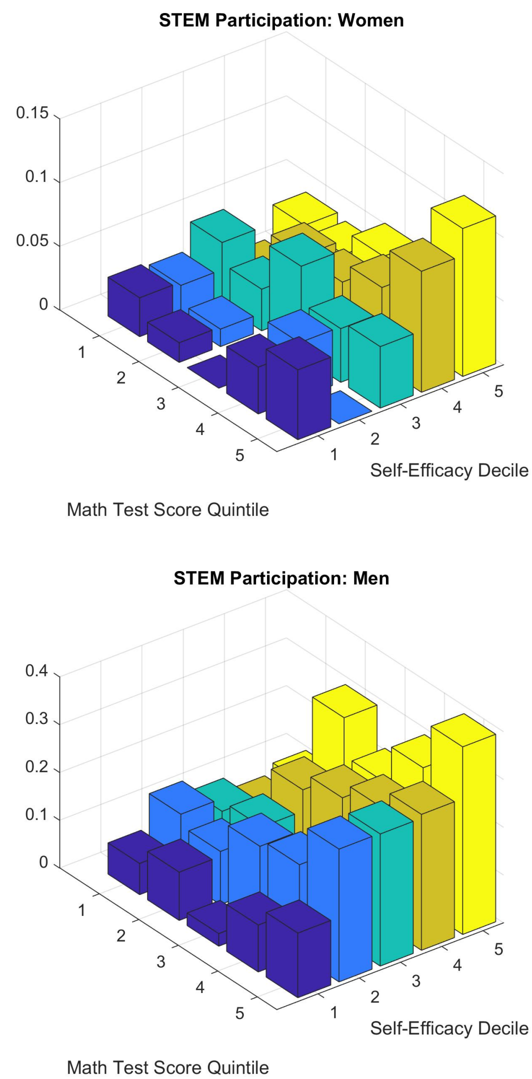
Note: Table C.1.8 examines the goodness of fit for labor market outcomes in the discrete choice model. Goodness of fit is tested using a t-test for employment and hourly wages where the Null Hypothesis is $Model = Data$. Note: * p<0.10, ** p<0.05, *** p<0.01. Observed majors at enrollment and graduation follow from the sample described in Section 4.2. Simulated results come from the 200,000 observations simulated for each gender using estimated model parameters.

**Figure C.1.1:** Marginal Distribution of Math Ability by Gender



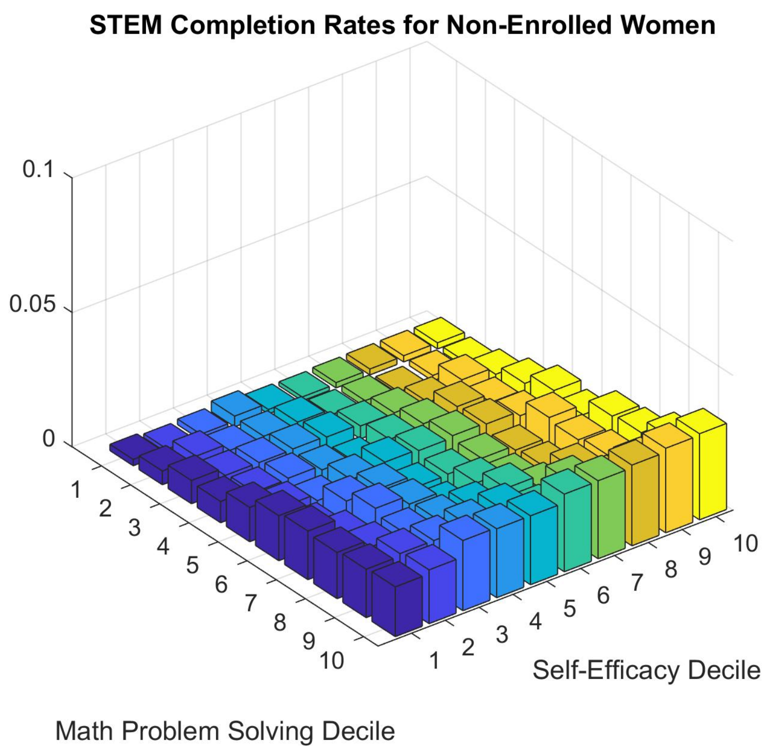

Source: Educational Longitudinal Study of 2002. Note: Figure C.1.1 shows the marginal density of the math problem solving and the self-efficacy components by gender.

**Figure C.1.2:** Sorting into STEM Enrollment by Baseline Test Scores



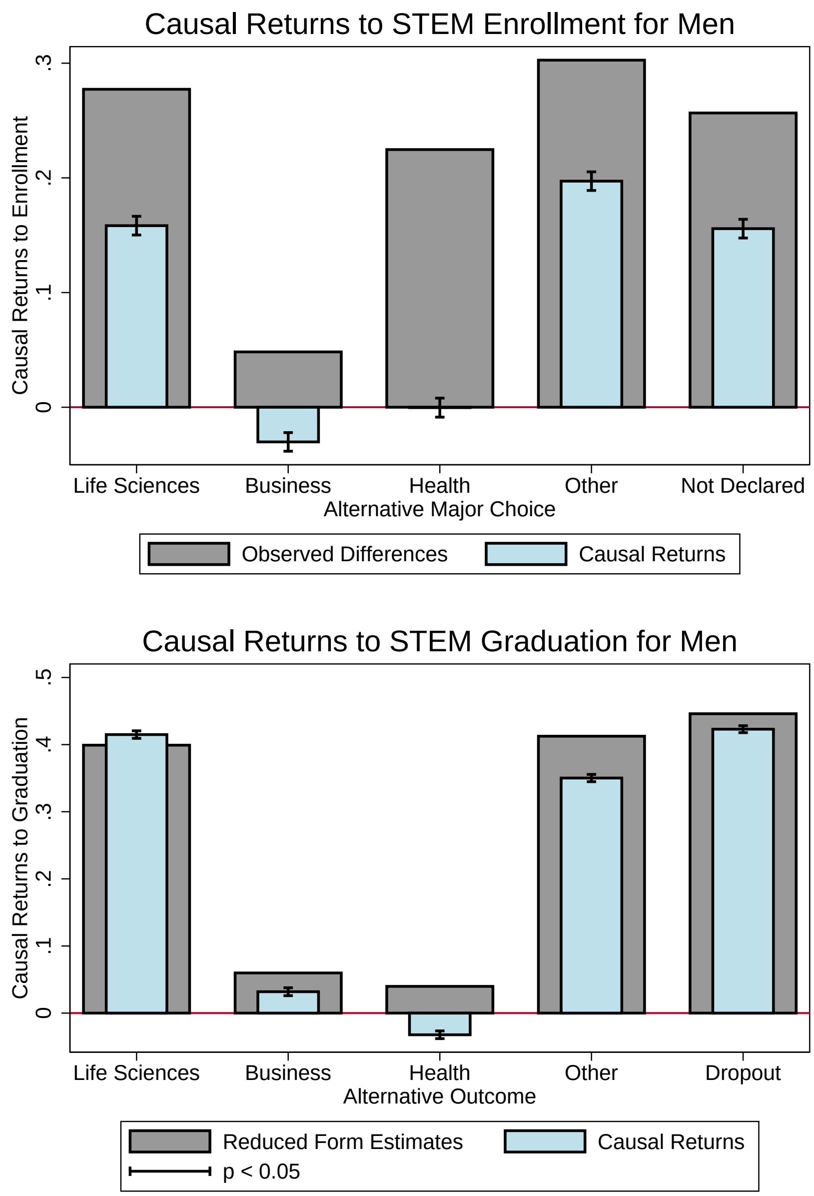

Source: Educational Longitudinal Study of 2002. Note: Figure C.1.2 explores how women and sort into STEM enrollment based on baseline math test scores and self-efficacy. These measures are classified into gender-specific quintiles.

**Figure C.1.3:** STEM Completion Rates for Women Enrolled in Other Majors



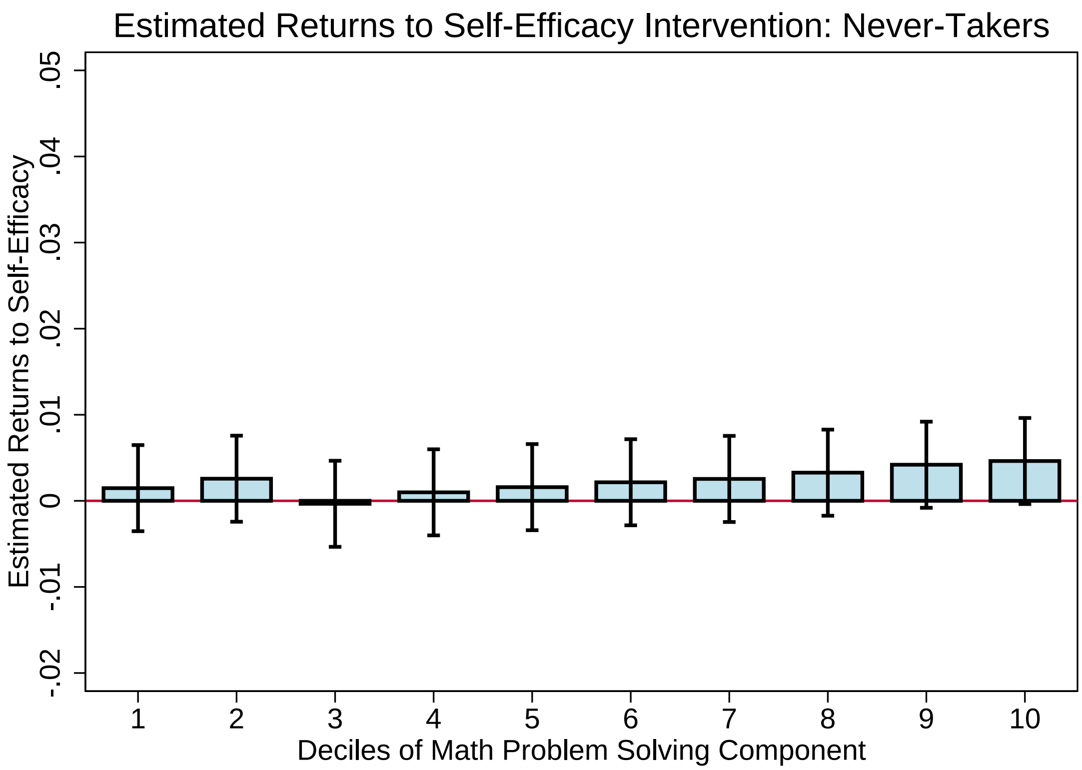

Source: Educational Longitudinal Study of 2002.
Note: Figure C.1.3 shows the share of women who graduate from a math-intensive major by age 26 among those who had not initially enrolled in these fields by the joint decile of the math problem solving and the self-efficacy ability components. The deciles of problem solving and self-efficacy are defined relative to the within-female ability distribution.

**Figure C.1.4:** Estimated Causal Returns to STEM Majors for Men




Source: Educational Longitudinal Study of 2002. Note: Figure C.1.4 presents the returns to enrollment in math-intensive STEM majors for men. The returns are estimated separately against each alternative major choice, $m_1 \in \mathcal{M}_\infty$ and educational outcome $m_2 \in \mathcal{M}_\in$ in each panel, respectively. The returns presented represent the average treatment effect (ATE) of each major, as defined in equation (4.13). The returns to enrollment and graduation are compared to the raw wage differences among STEM enrollees and completers, respectively, against the alternative outcome. The 'Causal Returns' estimate follows from estimated parameters in the dynamic discrete choice model.

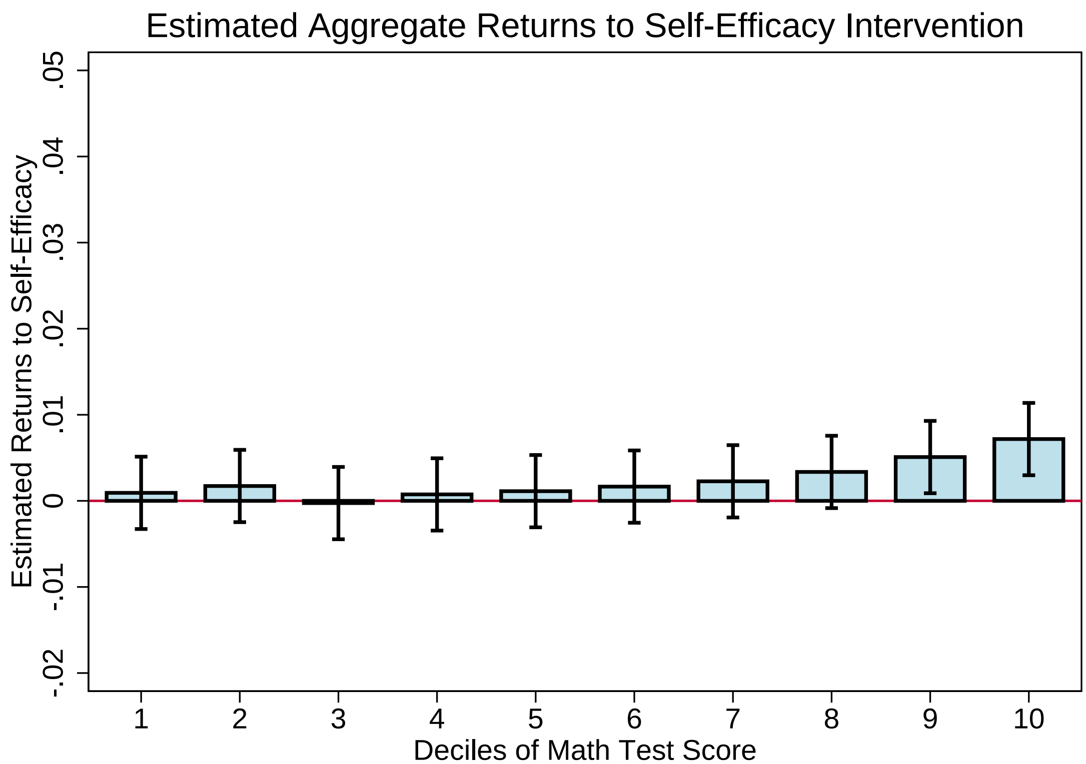
**Figure C.1.5:** Effects of Self-Efficacy Intervention on Wages for Never-Takers




Source: Educational Longitudinal Study of 2002.

Note: Figure C.1.5 presents the impact of a 0.25 SD boost in $\theta_{SE}$ for the set of female college students who do not start in a STEM field and do so under the intervention, as well. I present heterogeneous impacts across the $\theta_C$ distribution. The second panel presents the same analysis for compliers, that is, the students who in the baseline *did not* enroll in STEM, but did so following the intervention.

**Figure C.1.6:** Effects of Self-Efficacy Intervention on Wages by Test Score Decile




Source: Educational Longitudinal Study of 2002.
Note: Figure C.1.6 shows the impact of the 0.25 SD self-efficacy intervention across the baseline math test score deciles. This exercise corresponds to a policy intervention in which policymakers could only target students at a specific decile of the test score distribution.

## D.1 Identification of the Measurement System

This section presents the identification of the measurement system presented in Section 4.3. The identification of the distribution of unobserved ability follows the formal arguments presented in Carneiro et al. (2003), Hansen et al. (2004) and Heckman et al. (2006). In the measurement system presented in equations (4.7)-(4.10), the covariance between all test scores is observed and relied on as part of the identification strategy. Throughout this section, I keep the conditioning on $\boldsymbol{X}$ implicit. Let $C_j$ denote math test test score measures ($j = 1, 2, 3$). Using the covariances between these test scores, I can compute:

$$Cov(C_1, C_2) = \alpha_{C_1}\alpha_{C_2}\sigma^2_{\theta,C}$$

$$Cov(C_2, C_3) = \alpha_{C_2}\alpha_{C_3}\sigma^2_{\theta,C}$$

$$Cov(C_1, C_3) = \alpha_{C_1}\alpha_{C_3}\sigma^2_{\theta,C}$$

where $\sigma^2_{\theta,C}$ represents the variance of the problem solving factor. As noted in Section 4, by normalizing the loading associated with the baseline math test score ($\alpha^C_1 = 1$), I get a system with three equations and three unknowns. I can therefore identify the remaining three unknown parameters $\alpha_{C_2}, \alpha_{C_3}$, and $\sigma^2_{\theta,C}$.

I can similarly identify the loadings associated with the problem solving factor for the math GPA measures, self-efficacy measures and reading test scores using their respective covariances as follows:

$$Cov(C_1, G_1) = \alpha_{C_1}\alpha_{G_1}\sigma^2_{\theta,C}$$

$$Cov(C_1, SE_1) = \alpha_{C_1}\alpha_{SE_1}\sigma^2_{\theta,C}$$

$$Cov(C_1, R_1) = \alpha_{C_1}\alpha_{R_1}\sigma^2_{\theta,C}$$

Since I have already identified $\alpha_{C_1}$ and $\sigma^2_{\theta,C}$, I can identify the remaining loadings $\alpha_{G_1}, \alpha_{SE_1}$, and $\alpha_{R_1}$ from each equation presented above. A similar argument applies to the identification of the following loadings: $\alpha_{SE_2}$,$\alpha_{R_2}$, and $\alpha_{R_3}$.

To identify the variance of the self-efficacy factor and the self-efficacy loadings in the math GPA and self-efficacy measures, I follow a similar argument. Recall that so far I have assumed that the three components of ability are independent, such that $\theta_C \perp \theta_{SE} \perp \theta_R$. I relax this assumption later. The covariances between these measures are given by:

$$Cov(G_1, SE_1) = \alpha_{G_1}\alpha_{SE_1}\sigma^2_{\theta,C} + \gamma_{G_1}\gamma_{SE_1}\sigma^2_{\theta,SE}$$

$$Cov(G_1, SE_2) = \alpha_{G_1}\alpha_{SE_2}\sigma^2_{\theta,C} + \gamma_{G_1}\gamma_{SE_2}\sigma^2_{\theta,SE}$$

$$Cov(SE_1, SE_2) = \alpha_{SE_1}\alpha_{SE_2}\sigma^2_{\theta,C} + \gamma_{SE_1}\gamma_{SE_2}\sigma^2_{\theta,SE}$$

$\sigma^2_{\theta,SE}$ represents the variance of the self-efficacy factor. As with the problem solving factor, I normalize the loading associated with the baseline self-efficacy measure ($\gamma_{SE_1} = 1$) leaving a system with three equations with three unknowns, given that all the $\alpha$ loadings have been identified in the previous step along with $\sigma^2_{\theta,C}$. I can therefore identify the remaining three unknown parameters $\gamma_{G_1}, \gamma_{SE_2}$, and $\sigma^2_{\theta,SE}$.

Following this framework, I can identify the variance of the remaining component of ability, the reading factor, as well as the reading loadings in the reading/English test score equations. The covariances between these measures are given by:

$$Cov(R_1, R_2) = \alpha_{R_1}\alpha_{R_2}\sigma^2_{\theta,C} + \eta_{R_1}\eta_{R_2}\sigma^2_{\theta,R}$$

$$Cov(R_1, R_3) = \alpha_{R_1}\alpha_{R_3}\sigma^2_{\theta,C} + \eta_{R_1}\eta_{R_3}\sigma^2_{\theta,R}$$

$$Cov(R_2, R_3) = \alpha_{R_2}\alpha_{R_3}\sigma^2_{\theta,C} + \eta_{R_2}\eta_{R_3}\sigma^2_{\theta,R}$$

where $\sigma^2_{\theta,R}$ represents the variance of the self-efficacy factor. As the $\alpha_{R_j}$ and $\sigma^2_{\theta,C}$ components are already identified, the system above includes three equations and four unknowns. By normalizing the loading associated with the baseline reading test score ($\eta_{R_1} = 1$), I can identify the remaining loadings ($\eta_{R_2}$ and $\eta_{R_3}$), as well as the variance of the reading ability component, $\sigma^2_{\theta,R}$.

Having secured the identification of all the loadings and the variance of each component of latent ability, I apply the following transformation to the measurement system:[55]

$$\frac{C_j}{\alpha_{C_j}} = \theta_C + \frac{\varepsilon_{C_j}}{\alpha_{C_1}} \tag{20}$$

I can apply Kotlarski (1967)'s theorem to equation (20) to non-parametrically iden-

[55] I show the transformation for math test score measures $C_j$, but the argument applies across all test score equations.

tify the distributions of:

$$f_{\theta_C}(.), f_{\varepsilon_{C_j}}(.) \tag{21}$$

Applying the same argument to equations (4.7)-(4.10) identifies the distributions of:

$$f_{\theta_{SE}}(.), f_{\varepsilon_{G_1}}(.), f_{\varepsilon_{SE_n}}(.), f_{\theta_R}(.), f_{\varepsilon_{R_k}}(.) \tag{22}$$

**Correlated Factors: Identification**

To identify the correlation between unobserved abilities, I follow Heckman et al. (2016) and Prada and Urzúa (2017). Self-efficacy measures depend on both $\theta_C$ and $\theta_{SE}$. The correlation between both components is generated through the following linear association:

$$\theta_{SE} = \alpha_1 \theta_C + \theta_A \tag{23}$$

where $\theta_A$ is an auxiliary factor, independent of $\theta_C$ ($\theta_C \perp \theta_A$). For math GPA and self-efficacy measures, we can thus re-write the measurement system as:

$$G_1 = \alpha_{G_1}\theta_C + (\gamma_{G_1}\alpha_1)\theta_C + \gamma_{G_j}\theta_A + \varepsilon_{G_j}$$
$$G_1 = \beta_{G_1}\theta_C + \gamma_{G_1}\theta_A + \varepsilon_{G_j}$$

where $\beta_{G_1} = \alpha_{G_1} + \gamma_{G_1}\alpha_1$. Note that $\beta_{G_1}$ and $\gamma_{G_1}$ are identified through the arguments presented above. This argument similarly holds for the self-efficacy measures, yielding a system with three equations and four unknown parameters:

$$\beta_{G_1} = \alpha_{G_1} + \gamma_{G_1}\alpha_1$$

$$\beta_{SE_1} = \alpha_{SE_1} + \gamma_{SE_1}\alpha_1$$

$$\beta_{SE_2} = \alpha_{SE_2} + \gamma_{SE_2}\alpha_1$$

The four unknown parameters are $\alpha_{G_1}, \alpha_{SE_1}, \alpha_{SE_2}$, and $\alpha_1$, which denotes the correlation between $\theta_C$ and $\theta_{SE}$. As in Prada and Urzúa (2017), I apply an additional assumption, requiring the problem solving factor to affect the baseline self-efficacy measure only indirectly, through its correlation with the self-efficacy factor.[56] As a result, since $\alpha_{SE_1} = 0$, the remaining parameters in the system above are identified.

To identify the correlation between the problem solving component and the reading component, a similar argument follows. I again rely on auxiliary factor, positing the linear correlation between these two components:

$$\theta_R = \beta_1\theta_C + \theta_R \tag{24}$$

Following the same argument presented above yields a system of four equations and three unknown parameters, requiring an additional assumption to identify the correlation between $\theta_C$ and $\theta_R$. I assume that the problem solving factor affects the English GPA measure only indirectly, through its correlation with the reading

[56] I apply the normalization to the baseline self-efficacy measure as it has the lowest loading on the problem solving factor (Figure 3). The estimated correlations are robust to the choice of the other self-efficacy measure or to any math GPA measure.

factor, as this measure has the lowest loading on the problem solving factor. This assumption thus allows me to identify $\beta_1$ following the argument presented above. The results are robust to alternative assumptions regarding the indirect relationship between any reading/English measure and the problem solving component.

**Ability Updating Assumptions**

I note that if students' ability were to change between enrollment, dropout and final major decisions, the ability loadings ($\alpha$) in equations (4.2) and (4.3) would be biased, as the latent ability with which students sort into majors would be measured with error. The direction of the bias would directly depend on the structure of ability updating. While the structural literature cited above often assumes a linear updating process, where college ability is linear combination of pre-college ability and major-specific grades, the updating process could have different functional forms. As a result, signing the direction of the bias is not straight-forward. Furthermore, as Carroll et al. (1995) have argued, correcting for such measurement error is non-linear models, as in the multinomial probits used to estimate equations (4.2) and (4.3), is a more difficult problem than in linear models. Moreover, as there are no measures of math ability available in the follow-up surveys, I argue that pre-college ability remains a reasonable proxy for ability latent ability while in college and after graduation.

## D.2 Identification of Gender Differences in Latent Ability

In Section 4.3, I assumed that each component of latent ability for both males and females equaled zero. This assumption is required for the identification of the distribution of unobserved ability. However, given my interest in gender differences in latent math ability for understanding gaps in STEM participation and graduation, I relax this assumption. As a result, to identify gender differences in the means of unobserved abilities, I extend Urzua (2008)'s approach to identify these differences in a system in which observed measures depend on various latent factors.

Consider the math test score measure $C_J$ for men and women (I omit dependence on background characteristics for notational simplicity):

$$C_j^m = \varphi_j^m + \alpha_{C_j}^m \theta_C^m + \varepsilon_{C_j}^m$$
$$C_j^f = \varphi_j^f + \alpha_{C_j}^f \theta_C^f + \varepsilon_{C_j}^f$$

where $E(\varepsilon_{C_j}^m) = E(\varepsilon_{C_j}^f) = 0$. Let $\mu_C^h$ and $\mu_C^f$ denote the means of the distribution of latent math problem solving ability for males and females, respectively, and $\Delta_C$ represent the difference across genders, given by $\Delta_C = \mu_C^m - \mu_C^f$.[57] Assuming that $\varphi_j^m = \varphi_j^f$, equation (25) can then be re-written as:

$$[E(C_j^m) - E(C_j^f)] = \alpha_{C_j}^f \, \Delta_C - (\alpha_{C_j}^f - \alpha_{C_j}^m)\mu_C^f$$

[57]Similarly, $\Delta_{SE}$ represent the difference across genders in the self-efficacy component, given by $\Delta_{SE} = \mu_{SE}^m - \mu_{SE}^f$ and $\Delta_R$ captures differences in latent reading ability.

Assuming $\mu_C^f = 0$ normalizes the mean of the factor for females, though it could normalized to any number, making the assumption relatively innocuous. As such, gender differences in latent problem solving ability are given by:

$$[E(C_j^m) - E(C_j^f)] = \alpha_{C_j}^f \, \Delta_C \tag{25}$$

I apply the same analysis to math grades, self-efficacy measures and reading/English test scores, yielding the following equations: Equation (25) can then be re-written as:

$$[E(C_j^m) - E(C_j^f)] = \alpha_{C_j}^f \, \Delta_C \tag{26}$$

$$[E(G_1^m) - E(G_1^f)] = \alpha_{G_1}^f \, \Delta_C + \gamma_{G_1}^f \, \Delta_{SE} \tag{27}$$

$$[E(SE_n^m) - E(SE_n^f)] = \alpha_{SE_n}^f \, \Delta_C + \gamma_{SE_n}^f \, \Delta_{SE} \tag{28}$$

$$[E(R_k^m) - E(R_k^f)] = \alpha_{R_k}^f \, \Delta_C + \eta_{R_k}^f \, \Delta_R \tag{29}$$

The left-hand side can be directly computed for each of the nine test scores used in the measurement system. I note that while gender differences in math GPA, shown in equation (27) reflect both problem math solving ability and self-efficacy, as the variance of math GPA is largely explained by the latent problem solving factor (Figure 4.4), I assume that gender differences in math GPA reflect latent differences in problem solving ability, and are *not* reflective of gaps in latent math self-efficacy (note that the model still allows latent self-efficacy to affect math grades).[58] Given

[58]The results are not sensitive to this assumption. I find similar results in a measurement system in which observed self-efficacy measures are dedicated measures of latent self-efficacy, math

this set-up, I follow Urzua (2008) and identify gender differences in $\Delta_C$ from the average across all observed math test scores and grades affected by $\theta_C$.[59]

With $\Delta_C$ on hand, note that equation (27) has one unknown, $\Delta_{SE}$. As a result, $\Delta_{SE}$ is also identified from the average difference in $[E(SE_n^m) - E(SE_n^f)] - \alpha_{SE_n}^f \Delta_C$, weighted by the relative share of the variance of $\theta_{SE}$ explained by each of the two observed self-efficacy measures. The same logic applies to the identification of $\Delta_R$, which is computed from the weighted average gender difference in the reading SAT component, the English test score and the English GPA measure, given the prior identification of $\Delta_C$.

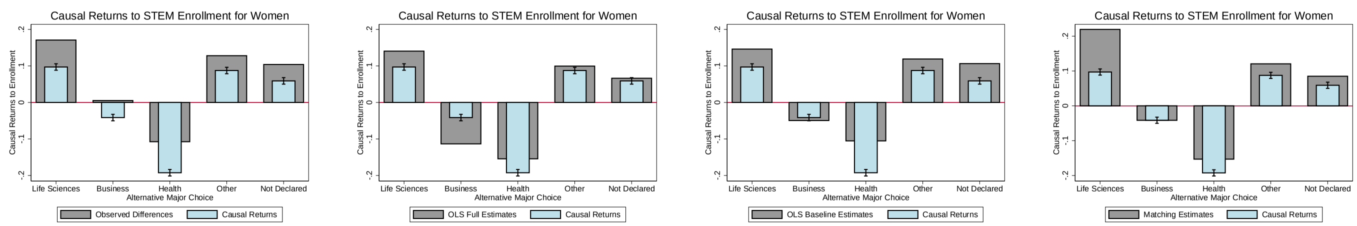
test scores depend on both math problem solving ability and self-efficacy and reading test scores depend on the three factors.

[59]Empirically, the average is calculated as a weighted average, where the weights are given by the relative share of the variance in test score measure $Cj$ or GPA $G_1$ explained by $\theta_C$. This procedure relaxes the linear average in gender differences in test scores in Urzua (2008).

## E.1 Reduced Form Returns to College Majors

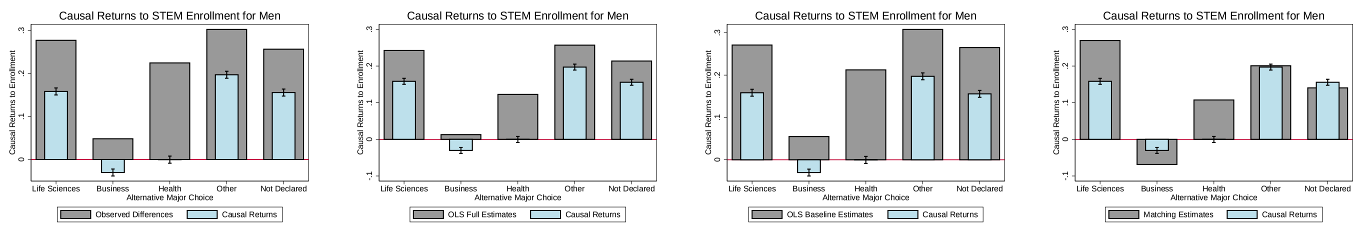
In this Section, I estimate the returns to STEM enrollment and graduation for men and women. I compare the model-based estimates of the average treatment effect presented in equation (4.13) to OLS regression and nearest-neighbor matching estimates.

**Figure E.1.1:** Reduced-Form Returns to STEM Enrollment for Women



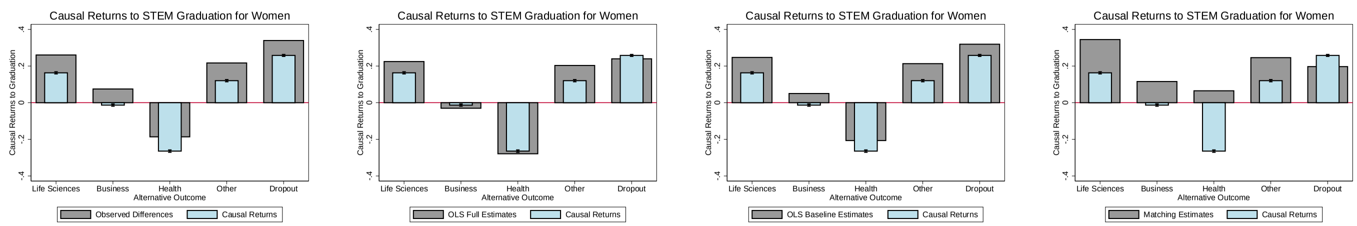

Source: Educational Longitudinal Study of 2002. Note: Figure E.1.1 shows the comparison of the estimated returns to STEM enrollment against reduced form estimates. The first panel analyzes observed differences across major alternatives. The second panel estimates an OLS regression with individual and family characteristics as control variables. The third panel includes baseline math, reading and self-efficacy test scores as control variables. The fourth panel presents estimates from nearest-neighbor matching including individual and family characteristics and test scores.

**Figure E.1.2:** Reduced-Form Returns to STEM Enrollment for Men




Source: Educational Longitudinal Study of 2002. Note: Figure E.1.2 shows the comparison of the estimated returns to STEM enrollment against reduced form estimates for men. The first panel analyzes observed differences across major alternatives. The second panel estimates an OLS regression with individual and family characteristics as control variables. The third panel includes baseline math, reading and self-efficacy test scores as control variables. The fourth panel presents estimates from nearest-neighbor matching including individual and family characteristics and test scores.

**Figure E.1.3:** Reduced-Form Returns to STEM Graduation for Women




Source: Educational Longitudinal Study of 2002. Note: Figure E.1.3 shows the comparison of the estimated returns to STEM enrollment against reduced form estimates. The first panel analyzes observed differences across major alternatives. The second panel estimates an OLS regression with individual and family characteristics as control variables. The third panel includes baseline math, reading and self-efficacy test scores as control variables. The fourth panel presents estimates from nearest-neighbor matching including individual and family characteristics and test scores.

## F.1 "Nudging" Policies

Could smaller-sized policies, such as "nudging" women towards STEM, increase participation rates? For instance, colleges could hire dedicated counselors to meet with women and discuss benefits associated with STEM while guiding them through first-year courses. In this context, it is also important to understand which majors these women would be coming from, as the benefits arising from STEM may be heterogeneous across different fields. The discrete choice model presented above allows me to identify the utility associated with each major (equations (4.1) and (4.3)), thus creating a cardinal ranking of all major choices for each student. As I calculate this utility for all individuals in the sample, I can identify those for whom the estimated utility between any two initial majors is largely equivalent, but who marginally choose a major not in a math-intensive field. This exercise is similar to the estimation of policy-relevant treatment effects, which first requires the identification of agents who would be affected by the policy of interest (Heckman et al., 2018; Humphries et al., 2017). "Nudged" individuals are defined by:

$$ID_{m_1} = 1\left[\sum_{m_j \in \mathcal{M}_1} V_{m_j} - V_{m_1} \leq \varepsilon\right] \;\; \forall\, m_j \;\in\; \mathcal{M}_1 \tag{30}$$

where $ID_{m_1}$ represents agents whose initial major is any of $m_j \in \mathcal{M}_1 \backslash m_1$ but who would be indifferent between having chosen major $m_1$ (in this case, a STEM field). $\varepsilon$ is an arbitrarily small neighborhood around the margin of indifference.[60] As the

[60] As in Heckman et al. (2018), I define the margin of indifference ($\varepsilon$) to be $\frac{V_{m_1}}{\sigma_{m_1}} \leq 0.01$, where $\sigma_{m_1}$ is the standard deviation of $V_{m_1}$.

agents identified in equation (30) are largely indifferent between their current majors towards math-intensive fields, I can examine how "nudging" these individuals towards STEM enrollment would affect aggregate enrollment rates. Furthermore, since I identify the distribution of $V_{m_j}$, I can discern the observed and unobserved characteristics of the agents included in the set $ID_{m_1}$. Heckman and Vytlacil (2007) and Heckman and Urzua (2010) note that in models with multiple choices, the indifference set may contain multiple margins. This analysis corresponds to identifying women who would be nudged from different majors into STEM, and analyzing the characteristics of agents in each subset of the indifference set $ID_{m_1}$.

I present the results in Table D1. I find that nudging women who were indifferent between their current choices and a math-intensive major would increase their aggregate enrollment in these fields from 4.4 percent to 4.7 percent. Women included in $ID_{m_1}$ are more likely to come from two-parent and higher income families compared to the rest of the sample. Furthermore, these women have higher endowments along the three dimensions of ability, surpassing the sample average in math problem solving by 0.33 standard deviations, in self-efficacy by 0.41 SDs and in the reading component by 0.14 SDs, as well. Lastly, over half of these women would be switching over from 'Other' majors, indicating that there is a margin for nudging high ability women towards enrolling in STEM and away from lower-paying fields.

In Table D1, I present the same results for women initially enrolled in STEM, who are largely indifferent between graduating from this field, yet choose to either switch fields or drop out of college. Women included in the set $ID_{m_1,G}$ represent upwards of 7 percent of initial STEM enrollees, indicating that a sizable share of women could be potentially nudged into remaining in STEM. A "nudging" policy at this stage could thus increase female graduation rates in this field from 43 percent to 50 percent, among initial enrollees. As with the "indifferent" individuals identified in Panel A, women in this indifference set have higher endowments in the three skill dimensions than women who are enrolled in STEM but end up not finishing this major.[61] The difference equals 0.17 standard deviations in the problem solving dimension, 0.20 SDs in self-efficacy and 0.15 in the reading component. The combination of these indifference sets suggests there is a margin for high ability women to be "nudged" into either enrolling or completing a STEM major. These policies could follow insights from Carrell et al. (2010) who find that a higher share of female faculty induces women towards STEM majors.

[61] I compare women in $ID_{m_1,G}$ to those who either switch majors or drop out to describe the types of individuals who could potentially be affected by a "nudging" policy.

**Table F.1.1:** Characteristics of Women Affected by "Nudging" policies

**Panel A.** Initial Major Choices

| | $ID_{m_1}$ | $ID_{m_1,m_2}$ | $ID_{m_1,m_3}$ | $ID_{m_1,m_4}$ | $ID_{m_1,m_5}$ | $ID_{m_1,m_6}$ |
|---|---|---|---|---|---|---|
| | Full Set | Life Sciences | Business | Health | Other | Not Declared |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Both Parents | 0.796 | 0.817 | 0.813 | 0.770 | 0.772 | 0.833 |
| Family Income | 11.21 | 11.49 | 10.82 | 10.85 | 11.19 | 11.41 |
| $\theta_C$ | 0.325 | 0.395 | 0.366 | 0.010 | 0.332 | 0.268 |
| $\theta_{SE}$ | 0.417 | 0.601 | 0.352 | 0.086 | 0.398 | 0.391 |
| $\theta_R$ | 0.137 | 0.212 | 0.006 | -0.221 | 0.176 | 0.108 |
| Share of Women | 0.0027 | 0.0004 | 0.0003 | 0.0001 | 0.0013 | 0.0004 |

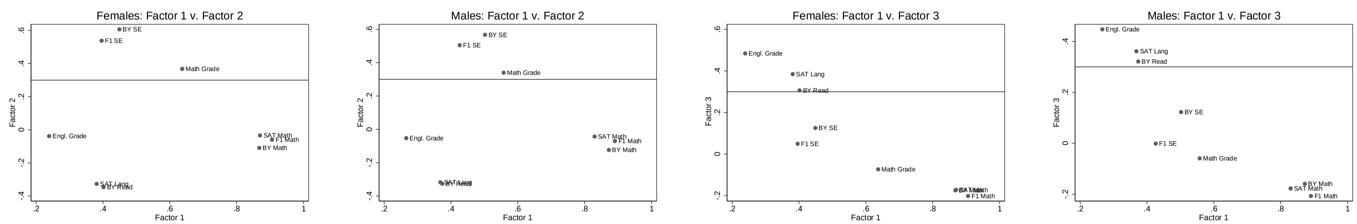
**Panel B.** Final Majors among STEM Enrollees

| | $ID_{m_1,G}$ | $ID_{m_1,G,S}$ | $ID_{m_1,G,D}$ |
|---|---|---|---|
| | Full Set | Major Switchers | Dropout |
| | (1) | (2) | (3) |
| Both Parents | 0.845 | 0.839 | 0.855 |
| Family Income | 11.30 | 11.32 | 11.26 |
| $\theta_C$ | 0.362 | 0.409 | 0.286 |
| $\theta_{SE}$ | 0.440 | 0.560 | 0.247 |
| $\theta_R$ | 0.157 | 0.173 | 0.133 |
| Share of Women | 0.076 | 0.045 | 0.031 |

Source: Educational Longitudinal Study of 2002.
Note: Appendix Table D1 displays the observed and unobserved characteristics who would be included in each respective nudging set. The first column in Panel A denotes the share of women in the full sample (0.27%) who would change their majors from their current choices to STEM, and the five columns on the right show which majors these women would be leaving behind. In Panel B, the "Share of Women" row is relative to the baseline STEM enrollment rate, such that among women currently not graduating from STEM, 7.6% would be included in the indifference set at graduation. $\theta_C$ represents the problem solving factor, $\theta_{SE}$ is the math self-efficacy component and $\theta_R$ is the reading ability component. Results are simulated from the estimates of the model.

## F.2 Exploratory Factor Analysis

**Figure F.2.1:** Loadings from Factor Analysis




Source: Educational Longitudinal Study of 2002. Note: Figure F.2.1 shows the loadings from exploratory factor analysis (EFA) using the nine observed measures included in the empirical analysis. As pointed out by citeprada2017one, any loading with a value greater than 0.3 is considered significant. EFA is done separately by gender and the panels reflect a comparison of the loadings in the first and second factor and the first and third factor, respectively. Engl. Grade refers to high school grades in English. Math Grade refers to high school grades in math courses. BY SE and F1 SE refer to the baseline and first follow-up self-efficacy measures, respectively. BY Math and F1 math refer to the baseline and first follow-up math test scores. SAT Math refers to the math SAT component, whereas SAT Read. does so for the reading portion of the test. BY Read considers the baseline English exam developed by ELS staff.

# Bibliography


Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*, 251–334.

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature 40*, 7–72.

Agan, A. (2013). The returns to community college. *Ph.D. Dissertation. The University of Chicago*.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics 11*(1, Part 1), 48–83.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ. 4*(1), 185–223.

Altonji, J. G. and N. Williams (2005). Do wages rise with job seniority? a reassessment. *ILR Review 58*, 370–397.

Anaya, L., F. Stafford, and G. Zamarro (2017). Gender gaps in math performance, perceived mathematical ability and college stem education: The role of parental occupation.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1-2), 343–375.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics 166*(1), 3–16.

Astorne-Figari, C. and J. D. Speer (2017). Are changes of major, major changes? the roles of grades, gender, and preferences in college major switching. *Unpublished Manuscript*.

Astorne-Figari, C. and J. D. Speer (2018). Drop out, switch majors, or persist? the contrasting gender gaps. *Economics Letters 164*, 82–85.

Avery, C. J. and T. J. Kane (2004). The boston coach program. *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*.

Bailey, M. J. and S. M. Dynarski (2011). Gains and gaps: Changing inequality in u.s. college entry and completion. *Whither Opportunity? Rising Inequality, Schools, and Childrens Life Chances*.

Beattie, G., J.-W. P. Laliberte, and P. Oreopoulo (2017). Thrivers and divers: Using non-academic measures to predict college success and failure. *Economics of Education Review*.

Beffy, M., D. Fougere, and A. Maurel (2012). Choosing the field of study in post-secondary education: Do expected earnings matter? *Review of Economics and Statistics 94*(1), 334–347.

Betz, N. E. and R. S. Schifano (2000). Evaluation of an intervention to increase realistic self-efficacy and interests in college women. *Journal of Vocational Behavior 56*(1), 35–52.

Black, D. and J. Smith (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics 24*(3), 701–728.

Black, D. A. and J. A. Smith (2004). How robust is the evidence on the effects of college quality? evidence from matching. *Journal of Econometrics 121*, 99–124.

Blau, F. D. and L. M. Kahn (2005). Do cognitive test scores explain higher u.s. wage inequality? *Review of Economics and Statistics 87*, 184–193.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of human Resources 43*(4), 972–1059.

Bound, J., M. F. Lovenheim, and S. Turner (2010). Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics 2*, 129–57.

Bowen, W., M. Chingos, and M. McPherson (2009). *Crossing the finish line: Completing college at America's public universities.* Princeton Univ Pr.

Brewer, D. J., E. R. Eide, and R. G. Ehrenberg (1999). Does it pay to attend an elite private college? cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources 34*.

Cameron, S. V. and J. J. Heckman (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of political Economy 109*(3), 455–499.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 46*, S13–S70.

Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics 20*(4).

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics 128*, 967–1015.

Carlana, M. (2018). Implicit stereotypes: Evidence from teachers' gender bias. *IZA DP No. 11659*.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. Technical report, National Bureau of Economic Research.

Carrell, S. and B. Sacerdote (2017). Why do college-going interventions work? *American Economic Journal: Applied Economics 9*, 124 – 151.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Carroll, R., D. Ruppert, and L. Stefanski (1995). Nonlinear measurement error models. *Monographs on Statistics and Applied Probability.(Chapman and Hall, New York) Volume 63*.

Cattan, S. (2013). Psychological traits and the gender wage gap. *The Institute for Fiscal Studies*.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research 1*(2), 245–276.

Cheng, A., K. Kopotic, and G. Zamarro (2017). Can parents growth mindset and role modelling address stem gender gaps?

Chetty, R. and N. Hendren (2018). The effects of neighborhoods on intergenerational mobility i: Childhood exposure effects. *Quarterly Journal of Economics*.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review 104*, 141–47.

Chevalier, A. (2011). Subject choice and earnings of uk graduates. *Economics of Education Review 30*(6), 1187–1201.

Conger, D. and M. C. Long (2010). Why are men falling behind? gender gaps in college performance and persistence. *The ANNALS of the American Academy of Political and Social Science*.

Cordero, E. D., S. H. Porter, T. Israel, and M. T. Brown (2010). Math and science pursuits: A self-efficacy intervention comparison study. *Journal of Career Assessment 18*(4), 362–375.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics 132*, 1593–1640.

Dickson, L. (2010). Race and gender differences in college major choice. *The Annals of the American Academy of Political and Social Science 627*(1), 108–124.

Dillon, E. W. and J. A. Smith (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics 35*, 45–66.

Dillon, E. W. and J. A. Smith (2018). The consequences of academic match between students and colleges. *NBER Working Paper No. 25069*.

Duckworth, A., C. Peterson, M. Matthews, and D. Kelly. *Personality Processes and Individual Differences- Grit: Perseverance amd Passion for Long-Term Goals.* American Psychology Association.

Eide, E. (1994). College major choice and changes in the gender wage gap. *Contemporary Economic Policy 12*(2), 55–64.

Ellison, G. and A. Swanson (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the american mathematics competitions. *Journal of Economic Perspectives 24*(2), 109–28.

EOP (2014). Commitments to action on college opportunity.

Fortin, N. M., P. Oreopoulos, and S. Phipps (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources 50*(3), 549–579.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics 34*.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104*(4), 1091–1119.

Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic perspectives 20*(4), 133–156.

Goodman, J., M. Hurwitz, and J. Smith (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics 35*, 829–867.

Grattoni, C. (2007). Spatial skills and mathematical problem solving ability on high school students. *Northwetern University*.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? testing measurement invariance using the confirmatory factor analysis framework. *Medical care 44*(11 Suppl 3), S78.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Diversity. culture, gender, and math. *Science (New York, NY) 320*(5880), 1164–1165.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018). Cyclical job ladders by firm size and firm wage. *Journal of Labor Economics 10*(2), 52–85.

Haltiwanger, J. C., H. R. Hyatt, and E. McEntarfer (2018). Who moves up the job ladder? *Journal of Labor Economics 36*(S1), S301–S336.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004). The effect of schooling and ability on achievement test scores. *Journal of econometrics 121*(1-2), 39–98.

Hastings, J., C. A. Neilson, and S. Zimmerman (2014). Are some degrees worth more than others? evidence from college admission cutoffs in chile. *NBER Working Paper No. 19241*.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016). Dynamic treatment effects. *Journal of econometrics 191*(2), 276–292.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–S246.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Heckman, J. J. and S. Urzua (2010). Comparing iv with structural models: What simple iv can and cannot identify. *Journal of Econometrics 156*(1), 27–37.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics 6*, 4875–5143.

Herrnstein, R. J. and C. Murray (1994). *Bell Curve: Intelligence and Class Structure in American Life*. Free Press.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics 91*, 717–724.

Hoxby, C. and C. Avery (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity 46*, 1–65.

Hoxby, C. M. and S. Turner (2015). What high-achieving low-income students know about college. *American Economic Review 105*, 514–17.

Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. *European journal of psychology of education 28*(1), 1–35.

Humphries, J. E., J. S. Joensen, and G. Veramendi (2017). College major choice: Sorting and differential returns to skills.

Humphries, J. E. and F. Kosse (2017). On the interpretation of non-cognitive skills. what is being measured and why it matters. *Journal of Economic Behavior and Organization 136*, 174–185.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Jiang, G. X. (2018). Planting the seeds for success: Why women in stem don't stick in the field.

Justman, M. and S. J. Méndez (2018). Gendered choices of stem subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of Education Review 64*, 282–297.

Kahn, S. and D. Ginther (2017). Women and stem. Technical report, National Bureau of Economic Research.

Katz, L. F., A. B. Krueger, et al. (1998). Computing inequality: have computers changed the labor market? *The Quarterly Journal of Economics 113*(4), 1169–1213.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *NBER Working Paper No. 20749*.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, National Bureau of Economic Research.

Kinsler, J. and R. Pavan (2011). Family income and higher education choices: The importance of accounting for college quality. *Journal of Human Capital 5*, 453 – 477.

Kinsler, J. and R. Pavan (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics 33*(4), 933–972.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics 20*(1), 69–76.

Krulik, S. and J. A. Rudnick (1989). *Problem Solving: A Handbook for Senior High School Teachers.* ERIC.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2017). Choice of majors: Are women really different from men? Technical report, National Bureau of Economic Research.

Light, A. and W. Strayer (2000). Determinants of college completion: School quality or student ability? *Journal of Human Resources 35*.

Lincove, J. A. and K. E. Cortes (2016). Match or mismatch? automatic admissions and college preferences of low-and high-income students. *NBER Working Paper No. 22559*.

Lindqvist, E. and R. Vestman (2011, January). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–28.

Lochner, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In *Handbook of the Economics of Education*, Volume 4, pp. 183–282. Elsevier.

Long, M. C. (2008). College quality and early adult outcomes. *Economics of Education Review 27*, 588–602.

Lovenheim, M. F. and C. L. Reynolds (2013). The effect of housing wealth on college choice: Evidence from the housing boom. *Journal of Human Resources 48*.

MacLeod, W. B., E. Riehl, J. E. Saavedra, and M. Urquiola (2017). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics 9*(3), 223–261.

Murphy, P. K. and P. A. Alexander (2000). A motivated exploration of motivation terminology. *Contemporary educational psychology 25*(1), 3–53.

NCES (2018). Graduation rates in usa: Fast facts. *Retrieved from NCES Website*.

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives 24*(2), 129–44.

Nix, S., L. Perez-Felkner, and K. Thomas (2015). Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among stem degree fields. *Frontiers in psychology 6*, 530.

Oreopoulos, P. and R. Ford (2016). Keeping college options open: A field experiment to help all high school seniors through the college application process. *NBER Working Paper No. 22320*.

Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics 33*, 493 – 520.

Perez-Felkner, L., S.-K. McDonald, B. Schneider, and E. Grogan (2012). Female and male adolescents' subjective orientations to mathematics and the influence of those orientations on postsecondary majors. *Developmental Psychology 48*(6), 1658.

Perez-Felkner, L., S. Nix, and K. Thomas (2017). Gendered pathways: How mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in psychology 8*, 386.

Pope, D. G. and J. R. Sydnor (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives 24*(2), 95–108.

Prada, M. F. (2014). Essays on the economics of ability, education and labor market outcomes. *Ph.D. Dissertation. University of Maryland College Park*.

Prada, M. F. and S. Urzúa (2017). One size does not fit all: Multiple dimensions of ability, college attendance, and earnings. *Journal of Labor Economics 35*(4), 953–991.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association 53*(284), 873–880.

Ransom, T. (2016). Selective migration, occupational choice, and the wage returns to college majors.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity? Rising Inequality, Schools, and Childrens Life Chances*.

Riegle-Crumb, C., B. King, E. Grodsky, and C. Muller (2012). The more things change, the more they stay the same? prior achievement fails to explain gender inequality in entry into stem college majors over time. *American Educational Research Journal 49*(6), 1048–1073.

Roderick, M., J. Nagoka, V. Coca, E. Moeller, K. Roddie, and D. Patton (2008). From high school to the future: Potholes on the road to college. *Consortium on Chicago School Research*.

Rodrıguez, J., F. Saltiel, and S. Urzúa (2017). Dynamic treatment effects of job training.

Rodriguez, J., S. Urzua, and L. Reyes (2016, apr). Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile. *Journal of Human Resources 51*(2), 416–460.

Rosen, S. (1981). The economics of superstars. *The American economic review 71*(5), 845–858.

Rothstein, J. M. (2004). College performance predictions and the sat. *Journal of Econometrics 121*(1-2), 297–317.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

Rumberger, R. W. and S. L. Thomas (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review 12*(1), 1–19.

Saltiel, F., M. Sarzosa, and S. Urzua (2017). Cognitive and socio-emotional ability. *Handbook on the Contemporary Economics of Education*.

Siegle, D. and D. B. McCoach (2007). Increasing student mathematics self-efficacy through teacher training. *Journal of Advanced Academics 18*(2), 278–312.

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

Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics 44*, 69–88.

Stinebrickner, R. and T. R. Stinebrickner (2003). Working during school and academic performance. *Journal of Labor Economics 21*, 449–472.

Stinebrickner, R. and T. R. Stinebrickner (2013). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies 81*(1), 426–472.

Topel, R. H. (1991). Specific capital, mobility, and wages: Wages rise with job seniority. *Journal of Political Economy 99*, 145–76.

Turner, S. E. and W. G. Bowen (1999). Choice of major: The changing (unchanging) gender gap. *ILR Review 52*(2), 289–313.

Urzua, S. (2008). Racial labor market gaps the role of abilities and schooling choices. *Journal of Human Resources 43*(4), 919–971.

Wiswall, M. and B. Zafar (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Wiswall, M. and B. Zafar (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Xie, Y. and K. A. Shauman (2003). *Women in Science*. Harvard University Press,.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*(3), 545–595.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics 32*, 711 – 754.

Zimmerman, S. D. (2017). Making the one percent: The role of elite universities and elite peers. *NBER Working Paper No. 22900*.