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

Fernando Saltiel*

Job Market Paper

Click Here for Latest Version

January 21, 2019

Abstract

This paper studies the relationship between pre-college skills and the gender gap in STEM majors. Using longitudinal data for the United States, I estimate a discrete choice model of initial and final major choices in which 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.

Keywords: Unobserved Heterogeneity, Major Choices, Gender Gaps, Treatment Effects.

*Fernando Saltiel. University of Maryland, Department of Economics, 3114 Tydings Hall, College Park MD, 20740. Email: saltiel@econ.umd.edu. I would like to thank Sergio Urzua for this invaluable guidance and support in this project. I am grateful to Melissa Kearney, Judy Hellerstein, Lesley Turner, Nolan Pope, Katharine Abraham, Cristian Sanchez, Camila Galindo-Pardo for their comments and suggestions. I am indebted to seminar participants at the University of Maryland and at the World Bank's Africa Region Gender Innovation Lab for helpful comments and feedback. All the data used in the paper was stored and managed on a secure server.

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 paper, 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] Other skill components, 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

[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.

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 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 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

[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.

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 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 grad-

[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.

[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).

uation 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 paper 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 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).

[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.

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 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 paper 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.

2 Data Sources and Summary Statistics

Data Sources

This paper 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 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 A.1, I show how the different restrictions result in the final sample used in the paper.

Summary Statistics

In Table 1, I present descriptive statistics on the sample used in this paper. 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] 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

[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.

[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).

[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

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 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 out-performing 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} grade survey. I complement this analysis by examining the relationship between math test scores and self-efficacy across the distribution in Figure 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 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 ma-

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.

jor, 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 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 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 A.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 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 3.

[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 A.3, I examine the contribution of baseline test scores to gender gaps in STEM participation. A reduced-form decomposition indicates these factors explain 15% of gaps at enrollment.

Lastly, in Table A.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.

3 Discrete Choice Model

This paper 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 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

[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 6.

[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.

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 $\left[D_{m_1}, D_{d_1}, D_{m_2}, D_g\right]$. 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^G_{i,m_1} depends on both observed characteristics and unobserved ability. It is given by:

$$V^G_{i,m_1} = \beta^G_{m_1} X^G_{i,m_1} + \alpha^G_{m_1} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,m_1} \quad for\ m_1 \in \mathcal{M}_1 \tag{1}$$

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

[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.

where X^G_{i,m_1} represents the vector of exogenous characteristics, $\boldsymbol{\theta_i}^G$ captures the vector of latent ability, and ε^G_{i,m_1} 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 $D^G_{i,m_1} = 1$ if $m_1 = \text{argmax}_{m_1 \in \mathcal{M}_1} \left\{ V^G_{i,m_1} \right\}$.

Final Major Choice

After initially enrolling in major D^G_{i,m_1}, 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^G_{i,d_1,m_1} represents an approximation to the net utility associated with continuing in four-year college after enrollment. V^G_{i,d_1,m_1} depends on observed and unobserved characteristics and is defined as follows:

$$V^G_{i,d_1,m_1} = \beta^G_{d_1,m_1} X^G_{i,d_1,m_1} + \alpha^G_{d_1,m_1} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,d_1,m_1} \tag{2}$$

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:

$$V^G_{i,m_2} = \beta^G_{m_2} X^G_{i,m_2} + \alpha^G_{m_2} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,m_2} \quad for\ m_2 \in \mathcal{M}_2 \tag{3}$$

X^G_{i,m_2} is a vector of exogenous characteristics, $\boldsymbol{\theta_i}^G$ represents latent ability endowments, and ε^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.

[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 paper 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 B.

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 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):

$$V^G_{i,g} = \beta^G_g X^G_{i,g} + \alpha^G_g \boldsymbol{\theta_i}^G + \varepsilon^G_{i,g} \tag{4}$$

As in the previous decision, $X^G_{i,g}$ is a vector of exogenous characteristics, $\boldsymbol{\theta_i}^G$ represents latent ability endowments, and $\varepsilon^G_{i,g}$ 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 $\left[D_{i,m_1}, D_{i,d_1}, D_{i,m_2}, D_{i,g}\right]$ 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 paper 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:

$$W^G_{i,s} = \beta^G_s X^G_{i,s} + \alpha^G_s \boldsymbol{\theta_i}^G + \varepsilon^G_{i,s} \;\; \forall \; s \; \in \; \mathcal{S} \tag{5}$$

where $X^G_{i,s}$ represents a vector of exogenous control variables determining hourly wages and $\varepsilon^G_{i,s}$ 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 par-

[24] As ELS data does not provide information on subsequent educational attainment, I classify students by their highest degree attained in the last survey round.

ticipate 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:

$$V^G_{i,se} = \beta^G_{se} X^G_{i,se} + \alpha^G_{se} \boldsymbol{\theta_i}^G + \varepsilon^G_{i,se} \;\; \forall \; s \;\in\; \mathcal{S} \tag{6}$$

where $X^G_{i,se}$ represents a vector of exogenous control variables determining hourly wages and $\varepsilon^G_{i,se}$ is the associated error term. Equations (1)-(6) imply that unobserved abilities $\boldsymbol{\theta}$ affect labor market productivity, initial and final major choices, and graduate school decisions. Equation (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 $\boldsymbol{\theta}$ 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 θ 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

[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 $\varepsilon_{i,m_1} \perp \varepsilon_{i,m_2} \perp \varepsilon_{i,s} \perp \varepsilon_{i,ke} \forall m_1 \in \mathcal{M}_1, \forall m_2 \in \mathcal{M}_2, s \in \mathcal{S}, \forall k \in \mathcal{S}$.

[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 θ to be fixed by the time of college enrollment, but not fixed from birth or indicative of gender differences in inherited ability.

(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.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 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 θ_C, math self-efficacy θ_{SE}, and reading ability θ_R.[28] Given the estimated loadings shown in Figure F.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 θ_{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:

$$C^G_{i,j} = \beta^G_{C_j} X^G_{i,C_j} + \alpha^G_{C_j} \theta^G_{C,i} + \varepsilon^G_{i,C_j} \tag{7}$$

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

$$G^G_{i,1} = \beta^G_{G_1} X^G_{i,G_1} + \gamma^G_{G_1} \theta^G_{SE,i} + \alpha^G_{G_1} \theta^G_{C,i} + \varepsilon^G_{i,G_1} \tag{8}$$

$$SE^G_{i,n} = \beta^G_{SE_n} X^G_{i,SE_n} + \gamma^G_{SE_n} \theta^G_{SE,i} + \alpha^G_{SE_n} \theta^G_{C,i} + \varepsilon^G_{i,SE_n} \tag{9}$$

[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.

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

$$R^G_{i,k} = \beta^G_{R_k} X^G_{i,R_k} + \eta^G_{R_k} \theta^G_{R,i} + \alpha^G_{R_k} \theta^G_{C,i} + \varepsilon^G_{i,R_k} \tag{10}$$

Across equations (7)-(10), $\boldsymbol{X}$ represents a vector of exogenous control variables and $\boldsymbol{\varepsilon}$ represents the error term. In Appendix B, I show how the measurement system secures the identification of the distribution of latent ability.[29] All error terms, ε_{i,C_j}, ε_{i,G_1}, ε_{i,SE_n}, and ε_{i,R_k} are mutually independent, independent of θ_C, θ_{SE}, θ_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 allow for the latent ability components to be correlated. In Appendix B, I discuss the assumptions required to identify the correlation between ability components. This paper 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 (7)-(10), which requires the normalization of one loading in each latent component and for $\boldsymbol{\theta}$ to be orthogonal to $\boldsymbol{X}$ and ε. The formal argument is laid out in Appendix B. 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

[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 θ_{SE} to affect performance in one math test score. I separately estimated a parallel measurement system allowing for θ_{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.

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 A.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 educational choice and outcome equations in the vector $\mathbf{X}_i$ and the vector of test scores in $\mathbf{T}_i$. Given the independence

[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).

assumptions invoked above, the likelihood for a set of I individuals is given by:

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

$$\mathcal{L} = \prod_{i \in I} \left[\iiint \prod_{k \in \mathcal{K}} f(T_{ik} \mid X_{ik}, \theta) \prod_{s \in \mathcal{S}} \left\{ Pr(D_{is} = 1 \mid \mathbf{X}_i, \theta) \left[f(Y_{is} \mid X_{is}, \theta) Pr(D_{ise} = 1 \mid X_{ise}, \theta) \right]^{D_{ise}} \right. \right.$$
$$\left. \left. [1 - Pr(D_{ise} = 1 \mid X_{ise}, \theta)]^{1 - D_{ise}} \right\}^{D_{is}} dF(\theta_C) dF(\theta_{SE}) dF(\theta_R) \right]$$

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 $(p_{1,k}, p_{2,k})$, 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,

[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.

[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).

yielding a total of 451,000 observations.

Table A.5 shows the variables used in the implementation of the model. These variables are the same for both genders. Meanwhile, Table A.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 A.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 χ^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 A.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 Model Results: Latent Ability

Measurement System

Tables A.9 and A.10 present the estimated coefficients from equations (7)-(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. 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 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.

[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.

Gender Differences in Latent Ability

In Section 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, 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 C, I present the formal 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, 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 A.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. 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 the distribution of women's self-efficacy is dominated by that of men.

Correlation of Latent Abilities

[35] As discussed in Appendix C, 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 shows these two measures loading on the problem solving factor as well. I do not focus much attention in Sections 5 and 6 exploring gaps in reading ability, however. These assumptions do not affect the interpretation nor the magnitude of subsequent empirical analysis.

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

5 Model Results: College Major Choices

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. Figure A.2 compares the estimated distributions of unobserved problem solving ability (Panel A) and self-efficacy (Panel B) across genders by initial major. 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 prob-

[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.

lem 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 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 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 5. Figure 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 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

[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.

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 A.11 and Figure A.3, 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.

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 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 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 A.4, 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 A.12, 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 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.

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, I examine whether closing gaps in math skills could lead to increased female enrollment in STEM. Following equation (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:

$$D^f_{m_1,\Delta} = g(X^f_{m_1}, \boldsymbol{\theta_C}^m, \boldsymbol{\theta_{SE}}^m) \quad \forall \;\; m_1 \in \mathcal{M}_1 \tag{11}$$

where $\boldsymbol{\theta_k}^m$ represents the male distribution of the k^{th} component of latent ability.[38] I present the results in Table A.13. 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 (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 A.13. Similar to the estimated impact on initial STEM enrollment, eliminating distributional gaps in the problem 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

[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.

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 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.

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]

[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.

[40]Jiang (2018) advances this literature by introducing a discrete choice model of college majors, where she finds

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:

$$W_{m_1} = D_{m_1,G}W_{m_1,G} + D_{m_1,S}W_{m_1,S} + D_{m_1,D}W_{m_1,D} \;\; \forall\, m_1 \; \in \mathcal{M}_1 \tag{12}$$

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:

$$ATE_{m_1,m_k} = E\left[W_{m_1} - W_{m_k}\right] \;\; \forall\, m_k \; \in \mathcal{M}_1 \tag{13}$$

As the discrete choice model presented in section 3 allows me to recover the latent wages across initial majors, I can estimate the gender-specific average returns to STEM enrollment from equation (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

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.

[41] In the context of the latent wage equation (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 ($\boldsymbol{\theta}$).

major m_k is given by:

$$\underbrace{E[Y_E|E = M_1] - [Y_E|E = M_k]}_{\text{Observed Difference}} = \underbrace{E[Y_{E=M_1} - Y_{E=M_k}]}_{\text{ATE (Enrollment)}} +$$
$$\underbrace{E[Y_{E=M_1} - Y_{E=M_k}|E = M_1] - E[Y_{E=M_1} - Y_{E=M_k}]}_{\text{Sorting Gains}} +$$
$$\underbrace{E[Y_{E=M_2}|E = M_1] - E[Y_{E=M_k}|E = M_k]}_{\text{Selection Bias}} \tag{14}$$

In Appendix D, 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 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 A.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 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.

[42]In Appendix D, 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 (13).

In Figure A.5, 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 just 44 percent of women finish this degree compared to 61 percent of men. On the other hand, following equations (5) and (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 on the untreated parameter (TUT). These param-

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

eters are defined as follows:

$$TT_{m_1,m_j} = E\left[W_{m_1} - W_{m_j} | D_{m_1} = 1\right] \quad \forall\ m_j\ \in \mathcal{M}_1 \tag{15}$$

$$TUT_{m_1,m_j} = E\left[W_{m_1} - W_{m_j} | D_{m_j} = 1\right] \quad \forall\ m_j\ \in \mathcal{M}_1 \tag{16}$$

In Table 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:

$$ATE_{m_1,S} = E\left[W_{m_1,G} - W_{m_1,S} | D_{m_1} = 1\right] \quad \forall\, m_1 \in \mathcal{M}_1 \tag{17}$$

$$ATE_{m_1,D} = E\left[W_{m_1,G} - W_{m_1,D} | D_{m_1} = 1\right] \quad \forall\, m_1 \in \mathcal{M}_1 \tag{18}$$

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 (17)) or dropping out (equation (18)).[44] I present the estimated conditional returns to STEM completion in Table 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 (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.

[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.

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, 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, I next examine whether skill-based interventions can offer a pathway for increasing women's STEM participation along with early-career labor market outcomes.

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 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

[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.

[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_τ) 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.

by:

$$
\begin{aligned}
\Delta^Y = E[Y(p') - Y] = \\
& E[Y(p') - Y|D_s(p') = 1, D_s = 0] \times \underbrace{P[D_s(p') = 1, D_s = 0]}_{STEM\ Enrollment\ Compliers} + \\
& E[Y(p') - Y|D_s(p') = 1, D_s = 1] \times \underbrace{P[D_s(p') = 1, D_s = 1]}_{STEM\ Enrollment\ Always-Takers} + \\
& E[Y(p') - Y|D_s(p') = 0, D_s = 0] \times \underbrace{P[D_s(p') = 0, D_s = 0]}_{STEM\ Enrollment\ Never-Takers}
\end{aligned} \tag{19}
$$

where D_s is a dummy variable which equals one for students enrolled in STEM. Equation (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 interventions for female college enrollees.[48]

Math Self-Efficacy Based Interventions

The results presented in Sections 5 and 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

[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.

[48] In Appendix E, 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 (1) and (3)), thus creating a cardinal ranking of all major choices for each student.

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 effects of simulated policies delivering math self-efficacy increases ranging from 0.1 to 1 standard deviations.[49] Despite the positive correlation between θ_{SE} and θ_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 (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:

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

In Table 7, I show that increasing women's math self-efficacy would increase STEM enrollment rates, as measured by the share of compliers. An 0.5 SD boost in θ_{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 θ_C distribution, depending on the complementarity of the two components of math ability in STEM fields. In the first panel of Figure 9, I examine how baseline participation rates change for women at each decile of the problem solving distribution. For women in the bottom θ_C decile, a 0.5 SD increase in self-efficacy would move

[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.

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 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 θ_{SE} boost of 0.25 standard deviations for the rest of the paper.[52]

In the second panel of Figure 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 participation 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.[53]

[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 θ_{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 θ_{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.

[53]The aggregate effect of the θ_{SE} interventions on STEM graduation rates (G) is thus given by:

$$\Delta^G = E[G(p') - G] = E[G(p') - G | D_s(p') = 1, D_s = 0] \times P[D_s(p') = 1, D_s = 0] +$$
$$E[G(p') - G | D_s(p') = 1, D_s = 1] \times P[D_s(p') = 1, D_s = 1] + E[G(p') - G | D_s(p') = 0, D_s = 0] \times P[D_s(p') = 0, D_s = 0]$$

I examine the effects of the intervention on STEM completion rates in the second panel of Table 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 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 10, I present the effects on always-takers, whose graduation rates increase largely uniformly across all θ_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 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 the third panel of Table 7, 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 θ_{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.[54]

For always-takers, on the other hand, the effect would be positive and significant, increasing hourly wages by almost 3 percent. In Figure 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 θ_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, as shown in Figure A.7. These results contribute to the nascent literature showing the productivity of non-cognitive skills in improving labor market 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

[54] In Figure A.7, I show 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 θ_{SE} boost.

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 7). 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.[55] 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 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 11, I expand upon these results by analyzing the heterogeneous benefits from the θ_{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 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

[55] These results match the heterogeneous returns to STEM by alternative major option shown in Table 5. Nonetheless, the wage returns to compliers capture a different parameter than the average treatment effects presented in equation (13).

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 A.8). 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.

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 paper, 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. 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.

References

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.

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.

Beffy, M., D. Fougere, and A. Maurel (2012). Choosing the field of study in postsecondary 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.

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.

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.

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. 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?

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

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.

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.

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.

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.

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.

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.

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. 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.

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.

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.

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 (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.

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.

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

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.

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. 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.

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.

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

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.

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

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.

9 Tables and Figures

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

Panel A. Latent Factors

Men					Women				
		Correlation					Correlation		
Factor	Mean (SD)	θ_C	θ_{SE}	θ_R	Factor	Mean (SD)	θ_C	θ_{SE}	θ_R
θ_C	0.166	1			θ_C	-0.001	1		
	(0.682)					(0.647)			
θ_{SE}	0.151	0.561	1		θ_{SE}	-0.001	0.471	1	
	(0.746)					(0.761)			
θ_R	-0.129	0.863	0.483	1	θ_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 displays the mean, standard deviation and correlation between the three ability components separately by gender. θ_C represents the problem solving factor, θ_{SE} is the math self-efficacy component and θ_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 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 θ_C)	0.070	-0.075	-0.180	0.055	0.040
	(0.006)***	(0.006)***	(0.006)***	(0.006)***	(0.006)***
ATE (High θ_C)	0.124	-0.008	-0.204	0.119	0.077
	(0.006)***	(0.006)	(0.006)***	(0.006)***	(0.006)***
ATE (Top θ_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 θ_{SE})	0.071	-0.071	-0.204	0.060	0.038
	(0.006)***	(0.006)***	(0.006)***	(0.006)***	(0.006)***
ATE (High θ_{SE})	0.123	-0.012	-0.180	0.115	0.079
	(0.006)***	(0.006)*	(0.006)***	(0.006)***	(0.006)***
ATE (Top θ_{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 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 (13). The treatment on the treated parameter (TT) refers to the parameter indicated in equation (15). The treatment on the untreated parameter (TUT) refers to the parameter indicated in equation (16). Lastly, the marginal treatment effect (MTE) is defined in equation (**??**). In Panel B, the ATE (Low θ_C) denotes the benefits for women below the math problem solving median, whereas the (High θ_C) refers to those above the median. A similar definition is used in Panel C.

Table 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 θ_C)	-0.063	0.236	0.446	0.370
	(0.026)***	(0.026)***	(0.018)***	(0.018)***
ATE (High θ_C)	0.118	0.383	0.440	0.342
	(0.016)***	(0.016)***	(0.012)***	(0.012)***
ATE (Top θ_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 θ_{SE})	0.071	0.291	0.439	0.355
	(0.027)***	(0.027)***	(0.018)***	(0.018)***
ATE (High θ_{SE})	0.066	0.360	0.442	0.346
	(0.016)***	(0.016)***	(0.012)***	(0.012)***
ATE (Top θ_{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 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. The ATE, TT, TUT and MTE parameters are as defined in equation (13)-(16). They are estimated from the estimated parameters in the discrete choice model. In Panel B, the ATE (Low θ_C) denotes the benefits for women below the math problem solving median, whereas the (High θ_C) refers to those above the median. A similar definition is used in Panel C.

Table 7: Aggregate Effects of Self-Efficacy Intervention

Panel A. Impact on Women's STEM Enrollment Rates by Intervention Sizes

	STEM Enrollment Rates
Baseline	0.045
+ 0.2 SD	0.050
	(0.001)***
+ 0.4 SD	0.056
	(0.001)***
+ 0.5 SD	0.060
	(0.001)***
+ 0.6 SD	0.062
	(0.001)***
+ 0.8 SD	0.069
	(0.001)***
+ 1 SD	0.077
	(0.001)***

Panel B. Impact on Graduation Rates and Hourly Wages

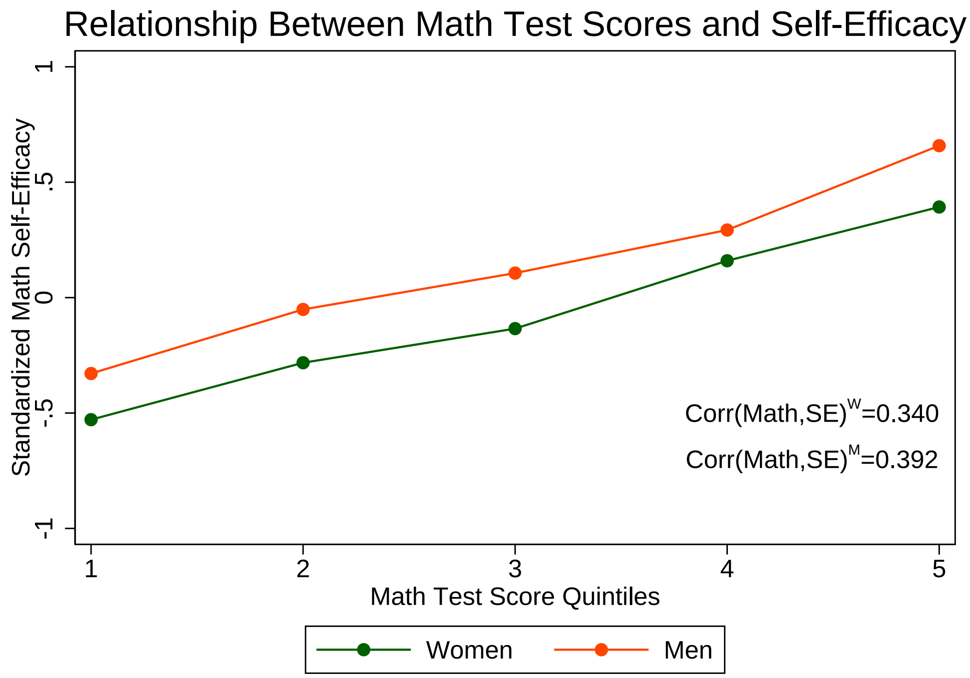
	Graduation Rates		Hourly Wages	
	Baseline	Post-Intervention	Baseline	Post-Intervention
	(1)	(2)	(3)	(4)
Full Sample	0.029	0.036	2.770	2.774
		(0.001)***		
Always-Takers	0.438	0.476	2.778	2.807
		(0.001)***		(0.019)*
Compliers	0.032	0.462	2.817	2.821
		(0.002)***		(0.043)
Never-Takers	0.010	0.011	2.769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

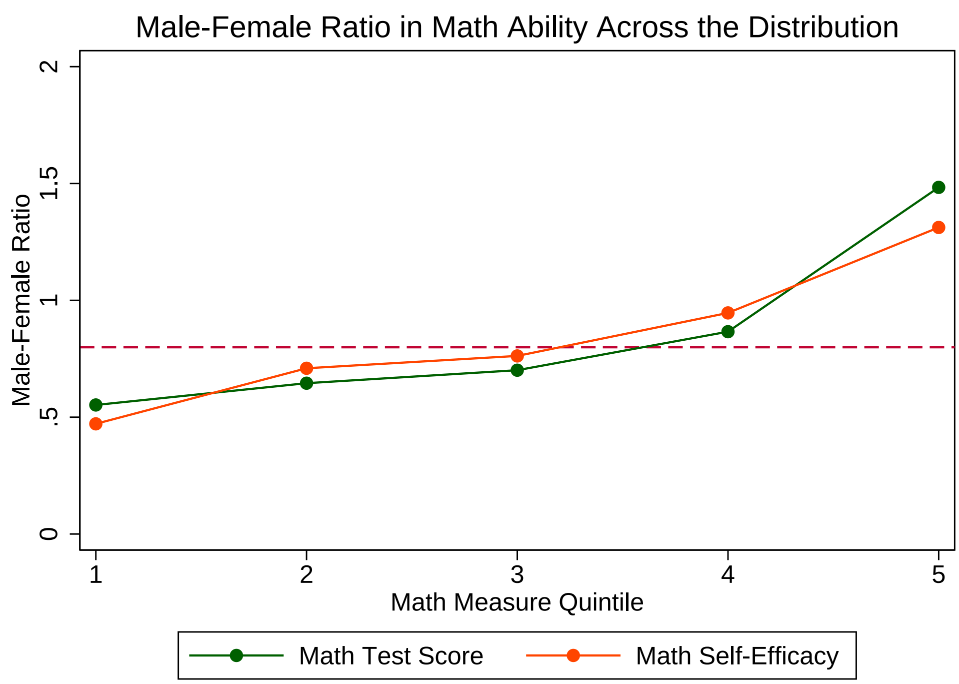
Panel C. Effects Across Complier Types

	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 7 examines the aggregate impact of the self-efficacy intervention on outcomes for female college students. The first panel analyzes interventions of varying magnitudes and their impact on enrollment rates. The standard errors refer to a comparison of enrollment rates relative to the baseline. The second panel explores the impact of the 0.25SD boost in θ_{SE} across the potential response groups and the SEs follow from a comparison of graduation rates relative to the baseline. The third panel 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.

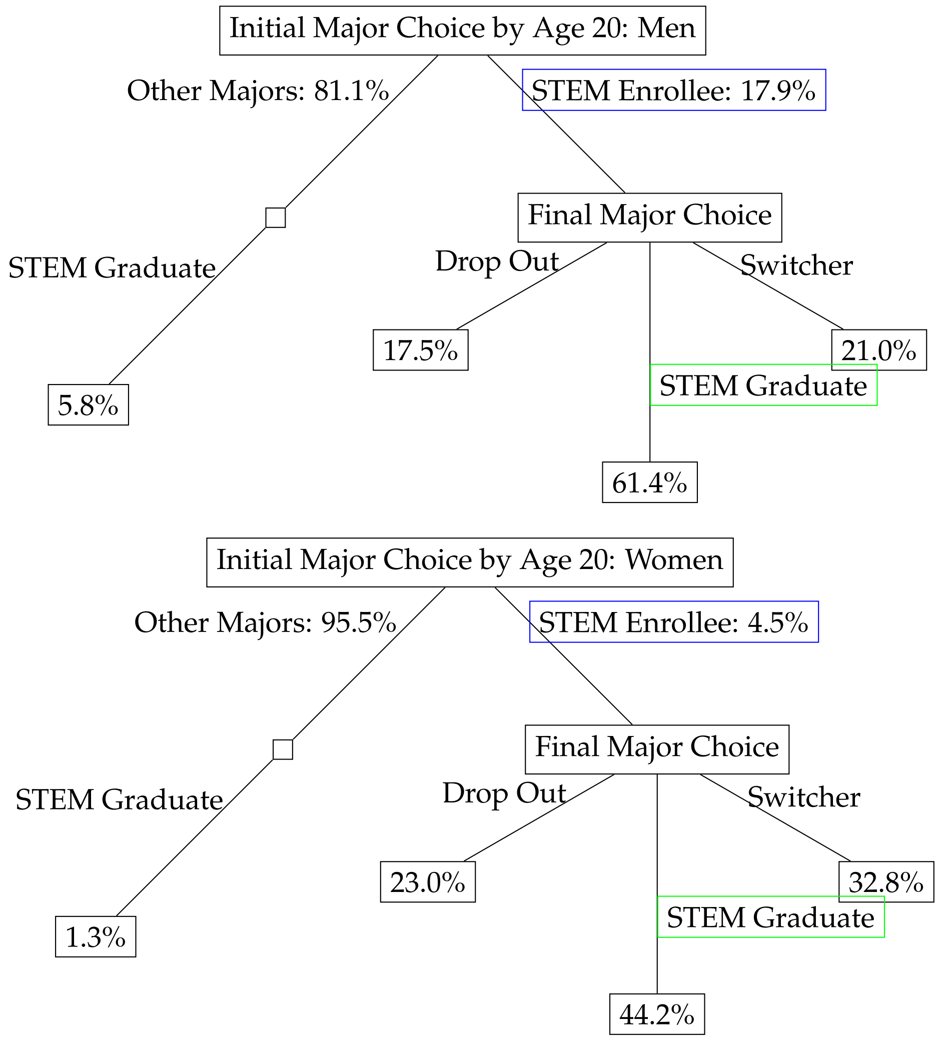
Figure 1: Math Test Score and Self-Efficacy Measures





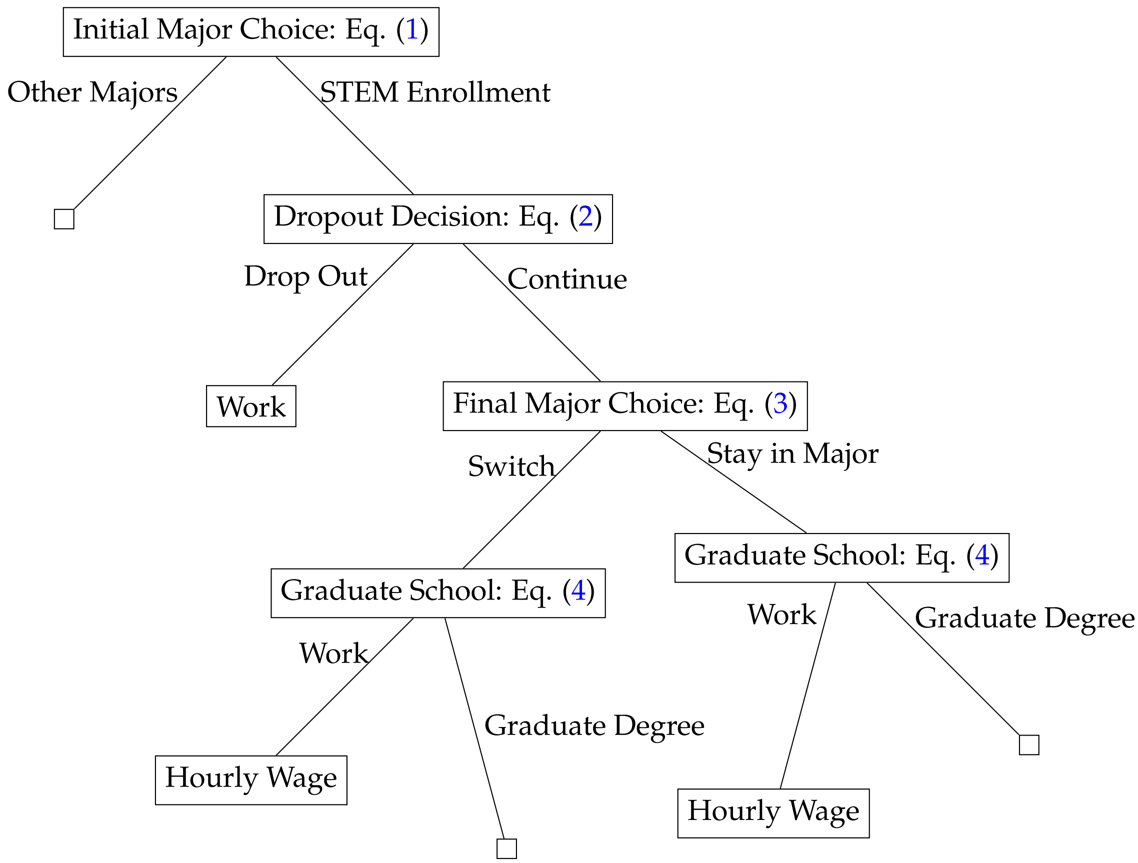
Source: Educational Longitudinal Study of 2002.
Note: This sample includes all students in the ELS 2002 who were enrolled in four-year college in the second follow-up survey. In Panel A, baseline self-efficacy scores are standardized normal (0,1), while math test scores are divided into quintiles. The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

Figure 2: STEM Enrollment Patterns and Subsequent Completion by Gender



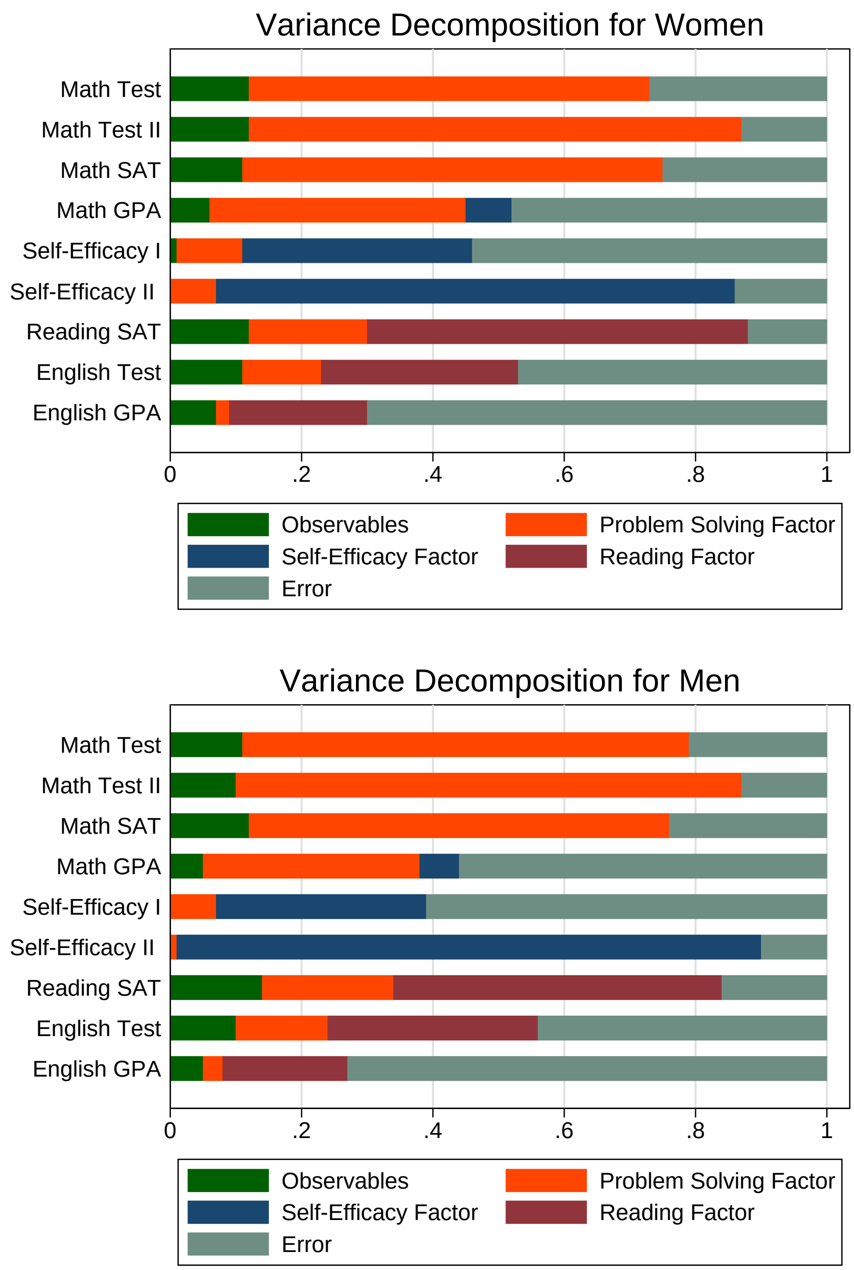
Source: Educational Longitudinal Study of 2002.
Note: This Figure depicts sequential progression through college majors for men and women by initial major choice (STEM and non-STEM). The left line depicts the share of students starting in other majors and the line below it, the share among these students 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 3: Structure of Discrete Choice Model



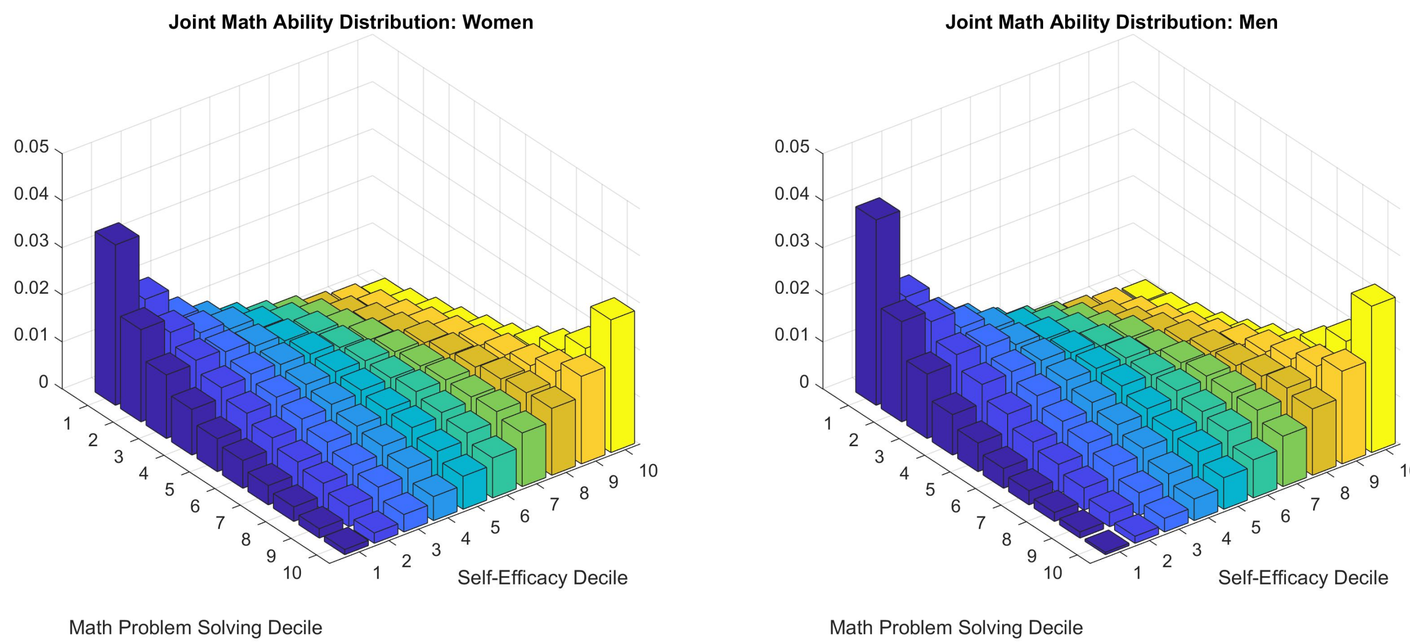
Source: Educational Longitudinal Study of 2002.
Note: This Figure depicts sequential progression through college majors for initial STEM enrollees. I note that despite this process not being depicted for students enrolled in 'Other' majors (due to limited space), college progression for these students,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

Figure 4: Measurement System: Variance Decomposition



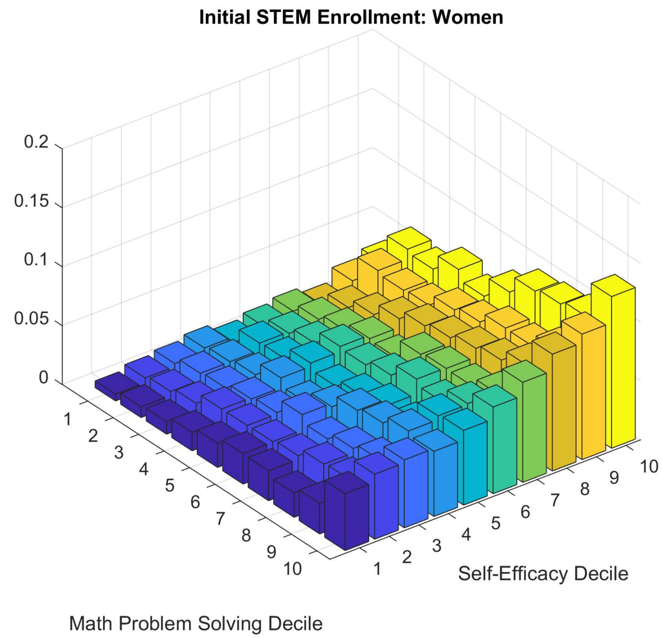
Source: Educational Longitudinal Study of 2002.
Note: This Figure shows the contribution of each variable to the variance of test scores using the simulated sample the model. The row Observables indicates the share of the variance of the measurement variables explained by the observed variables: child's 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.

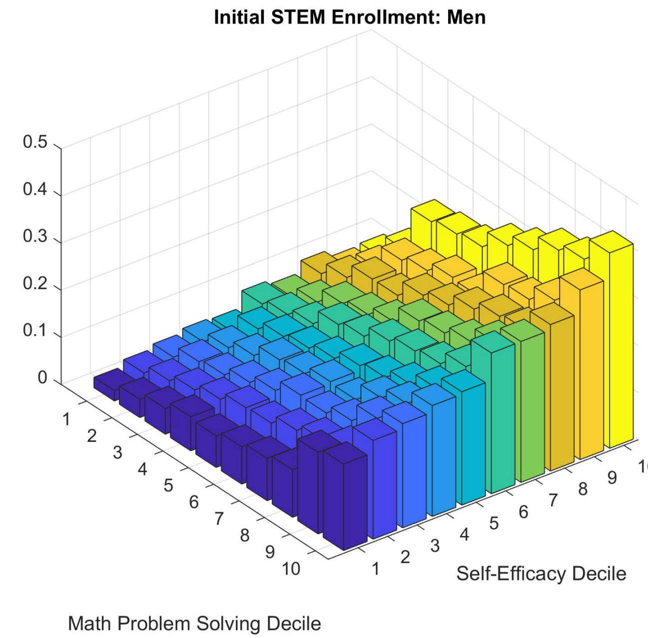
Figure 5: Joint Math Ability Density by Gender



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.

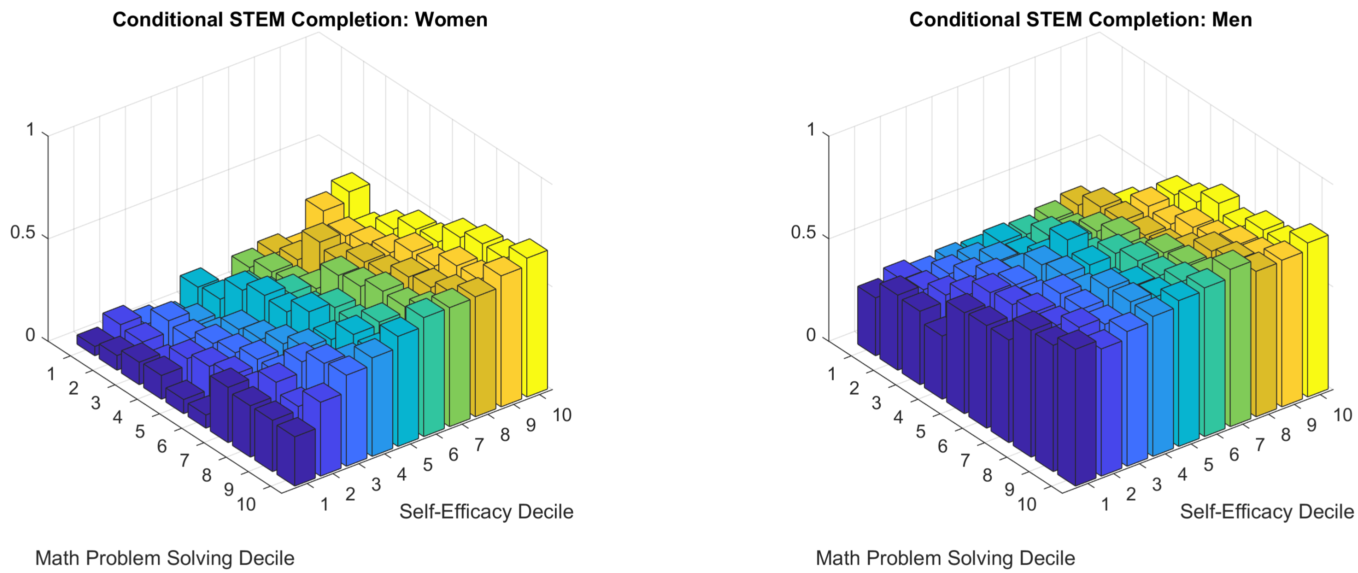
Figure 6: Share Initially Enrolled in a STEM Field by Gender





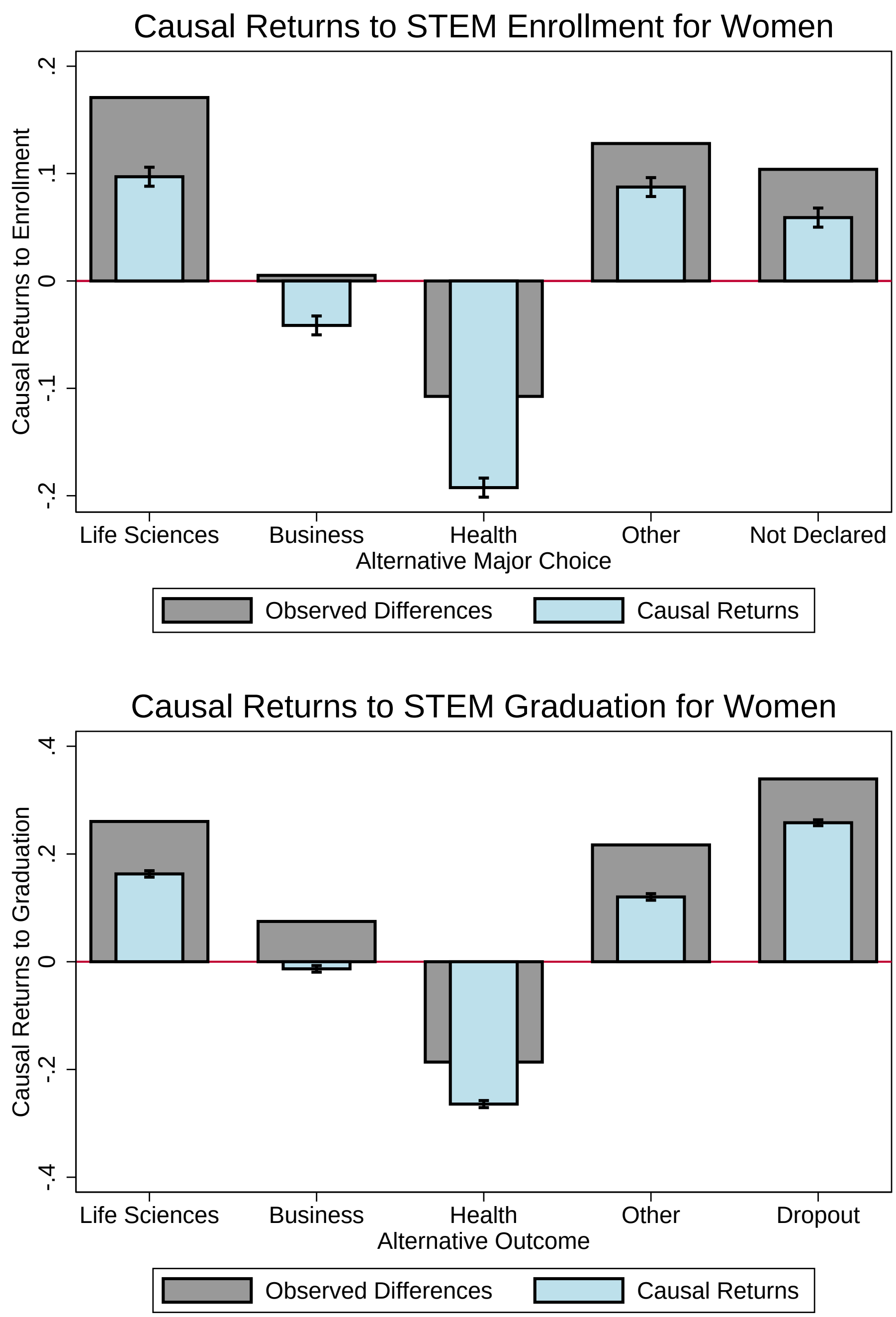
Source: Educational Longitudinal Study of 2002.
Note: Figure 6 shows the share of women and men initially enrolled in a math-intensive major in each joint decile of the math problem solving and the self-efficacy factors. The deciles of problem solving and self-efficacy are defined relative to the within-gender ability distribution. The results do not differ when defining ability deciles including both women and men.

Figure 7: Completion Rates for Students Initially Enrolled in STEM



Source: Educational Longitudinal Study of 2002.
Note: Figure 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.

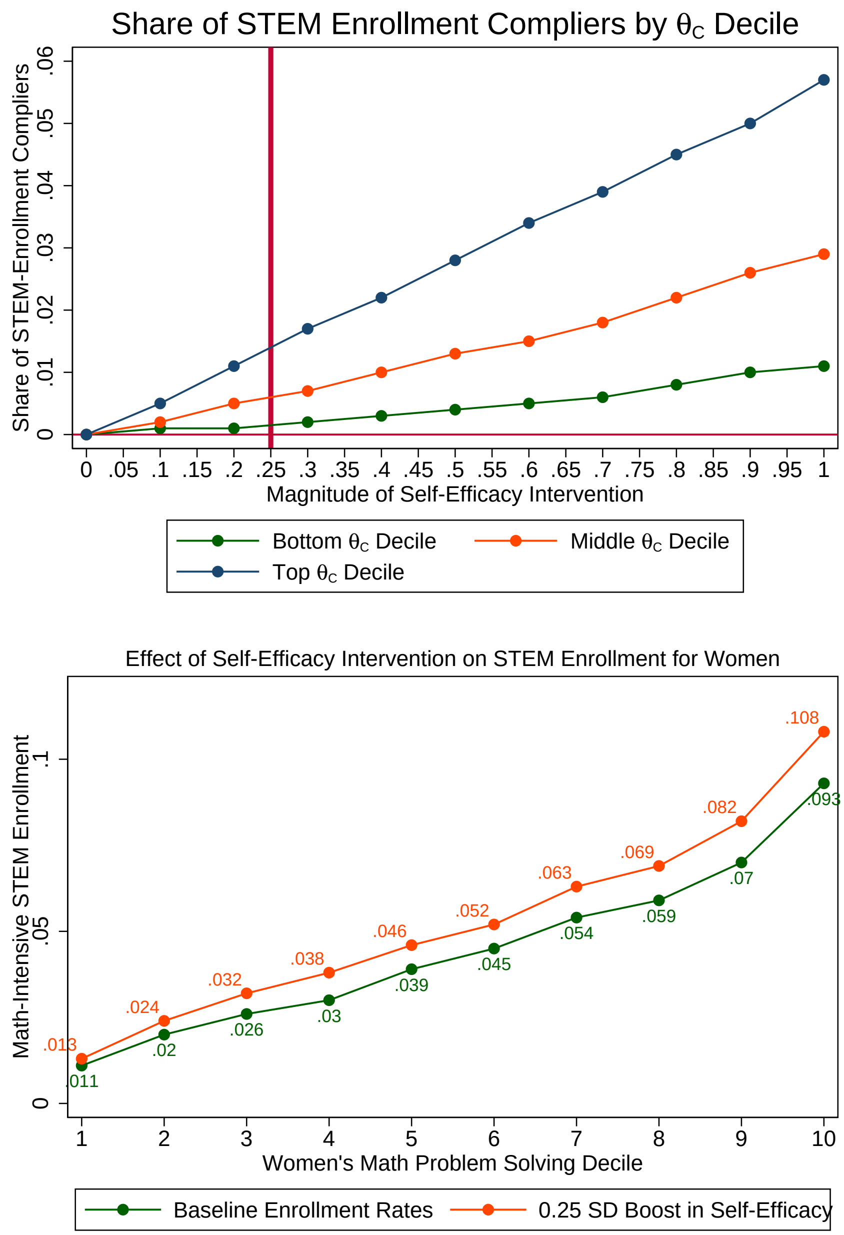
Figure 8: Estimated Causal Returns to STEM Majors for Women



Source: Educational Longitudinal Study of 2002.

Note: Figure 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 (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.

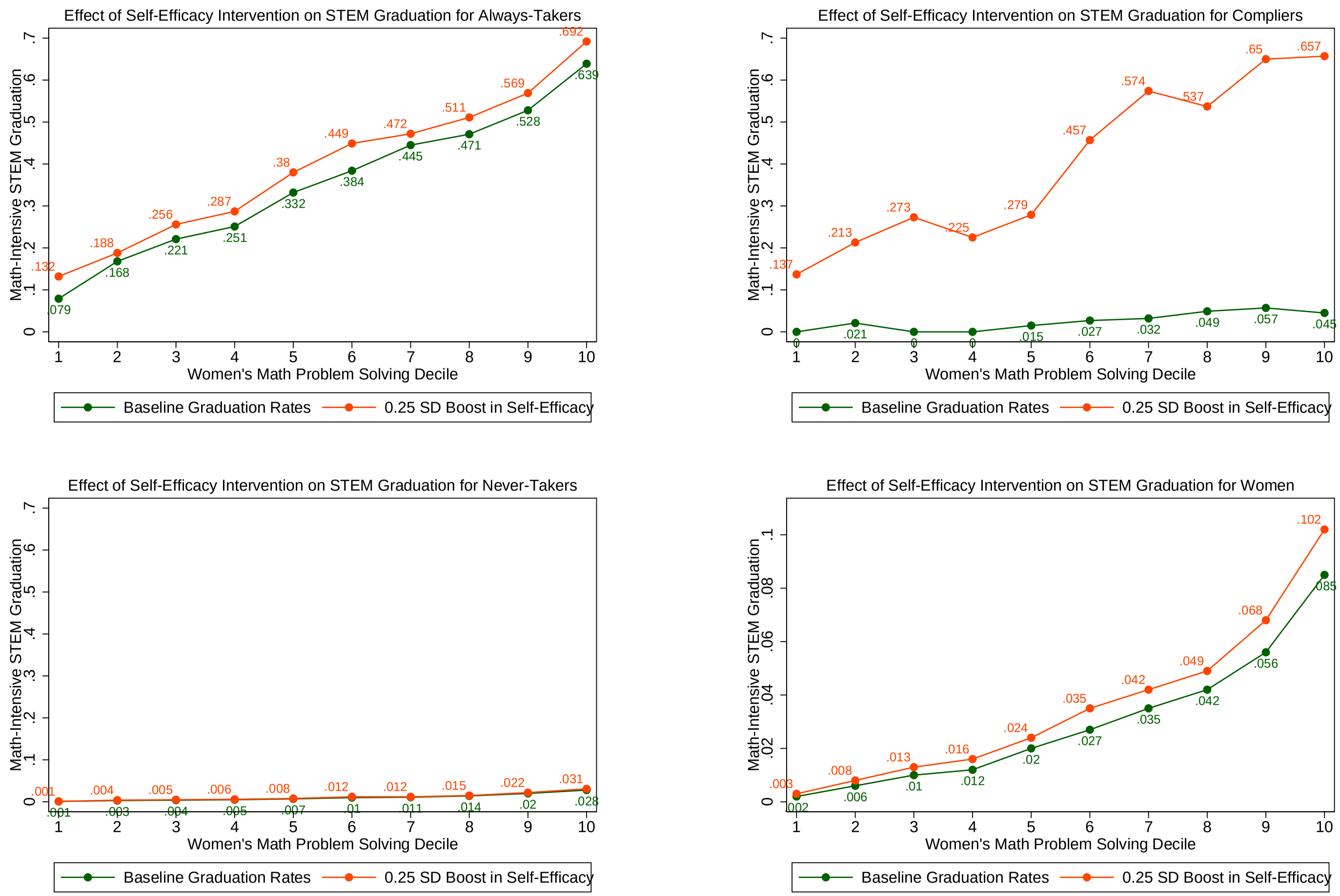
Figure 9: Estimated Impacts of Self-Efficacy Interventions on STEM Enrollment



Source: Educational Longitudinal Study of 2002.

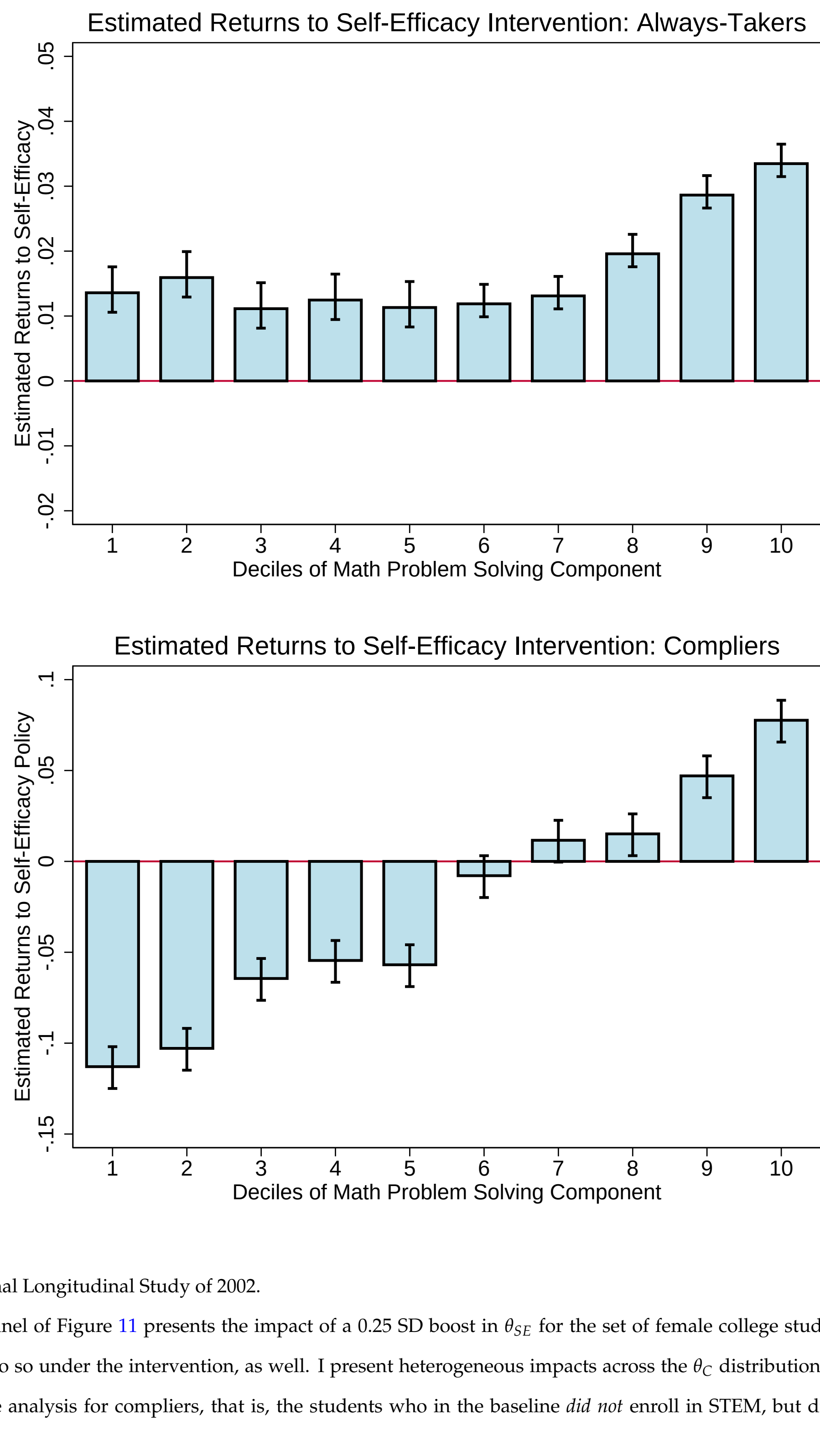
Note: The first panel of Figure 9 examines the impacts of self-efficacy interventions of varying magnitudes (from 0.1 SD to 1 SD) on women's enrollment rates in STEM, as measured by the share of compliers at each decile of the θ_C distribution. The vertical red line is drawn at 0.25 SDs, which is the magnitude of the policy intervention analyzed in Section 7. The second panel analyzes the impact of the 0.25 SD intervention on aggregate enrollment rates across each decile of the θ_C distribution. The green line captures baseline enrollment rates, whereas the orange line analyzes enrollment rates after the intervention.

Figure 10: Effects of Self-Efficacy Intervention on STEM Graduation by Potential Response Group



Source: Educational Longitudinal Study of 2002.
Note: Figure 10 presents the heterogeneous impacts of the 0.25 SD self-efficacy intervention on women's STEM graduation rates across the θ_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 θ_C distribution by combining the effect across the three potential response groups shown in the other three panels.

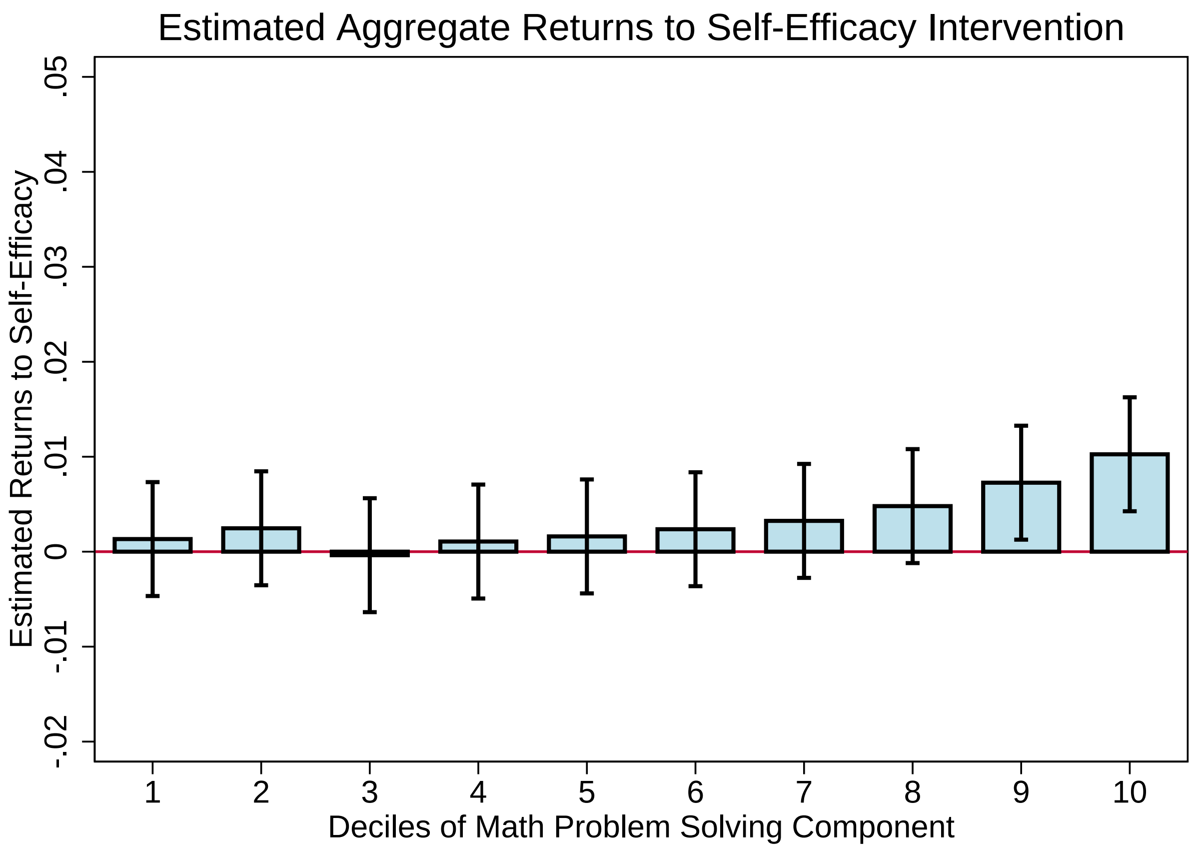
Figure 11: Effect of Self-Efficacy Intervention on Wages by Response Type



Source: Educational Longitudinal Study of 2002.

Note: The first panel of Figure 11 presents the impact of a 0.25 SD boost in θ_{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 θ_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 12: Estimated Impacts of Self-Efficacy Intervention on Hourly Wages



Source: Educational Longitudinal Study of 2002.

Note: Figure 12 presents the impact of a 0.25 SD boost in θ_{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 θ_C distribution.

Appendices

A Appendix Tables

Table A.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 A.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 A.2: Summary Statistics by Major

Panel A. STEM Enrollment Patterns by Gender

	Women			Men		
	STEM	Non-STEM	Difference	STEM	Non-STEM	Difference
	(1)	(2)	(3)	(4)	(5)	(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	Other	Dropout	STEM	Other	Dropout
	(1)	(2)	(3)	(4)	(5)	(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	Other	Dropout	STEM	Other	Dropout
	(1)	(2)	(3)	(4)	(5)	(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 A.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 A.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 $\times$ 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 A.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 A.4: Hourly Wages by Major

	Initial Major			Final Major		
Hourly Wages	Women (1)	Men (2)	Difference (3)	Women (4)	Men (5)	Difference (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 A.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 A.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 A.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	8.36
	(0.42)	(1.65)	(0.82)	(2.46)	(2.79)	(2.79)
White	-0.08	0.53	0.13	0.02	-0.08	-0.08
	(0.12)	(0.26)	(0.21)	(0.44)	(0.68)	(0.68)
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)					
θ_C	0.42	0.83	0.30	0.70	-0.40	-0.40
	(0.07)	(0.19)	(0.18)	(0.37)	(0.50)	(0.50)
θ_{SE}	0.43	0.59	0.39	0.11	-4.06	-4.06
	(0.11)	(0.26)	(0.20)	(0.37)	(1.55)	(1.55)
θ_R	-0.14	-0.55	0.20	-0.68	-0.94	-0.94
	(0.14)	(0.34)	(0.31)	(0.65)	(1.07)	(1.07)
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	3.83
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)

Source: Educational Longitudinal Study of 2002.
Note: Table A.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 A.7: Goodness of Fit: Educational Choices

Panel A. Initial Enrollment for Men and Women

	Women		Men	
	Observed	Simulated	Observed	Simulated
	(1)	(2)	(3)	(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	Simulated	Observed	Simulated
	(1)	(2)	(3)	(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	Simulated	Observed	Simulated
	(1)	(2)	(3)	(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 A.7 examines goodness of fit of educational decisions in the discrete choice model. Goodness of fit is tested using a χ^2 test where the Null Hypothesis is $Model = Data$. Observed majors at enrollment and graduation follow from the sample described in Section 2. Simulated results come from the 200,000 observations simulated for each gender using estimated model parameters.

Table A.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 A.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 2. Simulated results come from the 200,000 observations simulated for each gender using estimated model parameters.

Table A.9: 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)
θ_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)
θ_{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)
θ_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)

Source: Educational Longitudinal Study of 2002.
Note: Table A.9 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 A.10: 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)
θ_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)
θ_{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)
θ_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 A.10 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 A.11: 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 A.12: 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)
θ_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)
θ_{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)
θ_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 A.12 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 A.13: Sorting into College Continuation by Initial Major and Gender

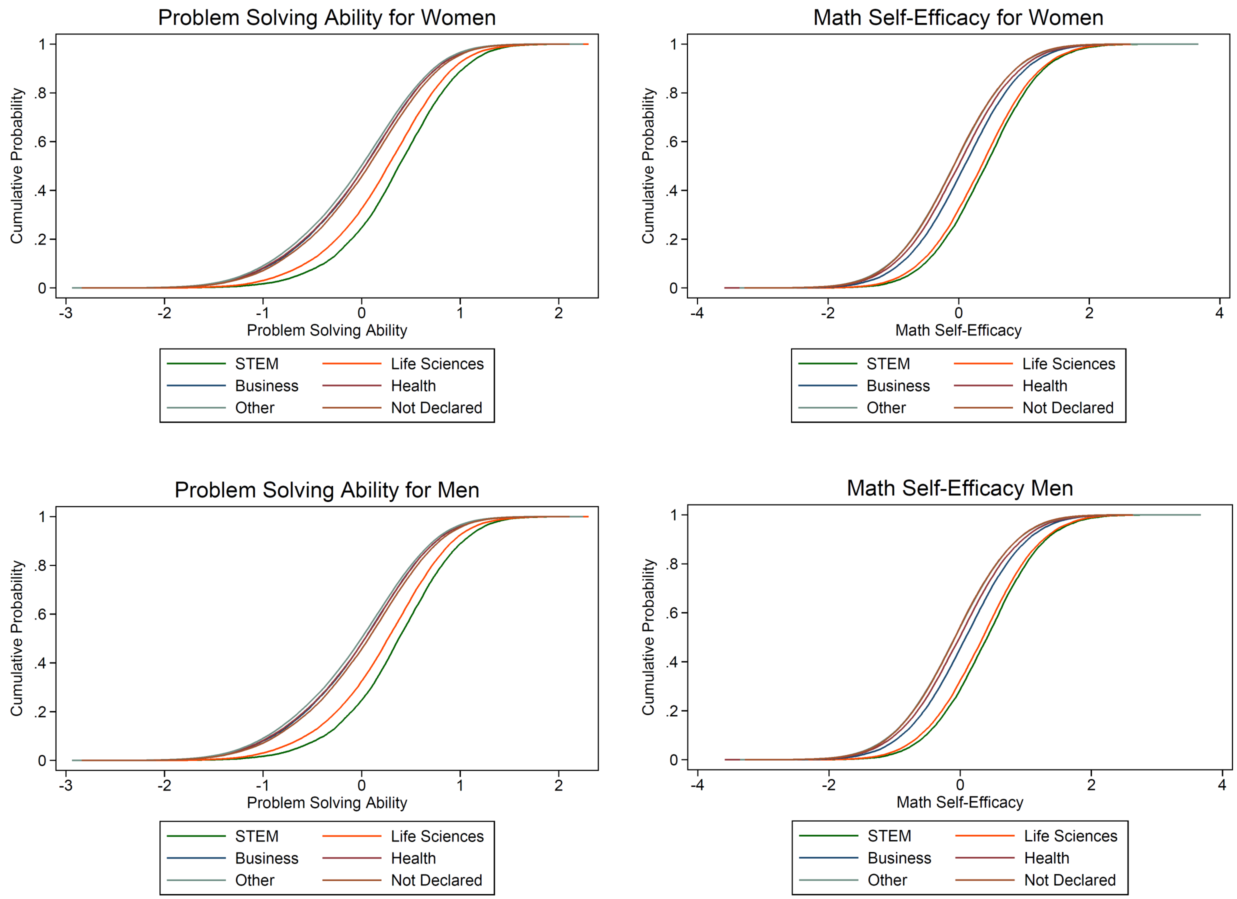
	STEM Enrollment Rates				
	Baseline	Problem Solving Comp.	Self-Efficacy Comp.	Joint Compensation	Male Share
	(1)	(2)	(3)	(4)	(5)
	0.045	0.053	0.053	0.062	0.179
	(0.001)	(0.001)***	(0.001)***	(0.001)***	
	STEM Graduation Rates				
	Baseline	Problem Solving Comp.	Self-Efficacy Comp.	Joint Compensation	Male Share
	(1)	(2)	(3)	(4)	(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 A.13 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 A.1: Marginal Distribution of Math Ability by Gender

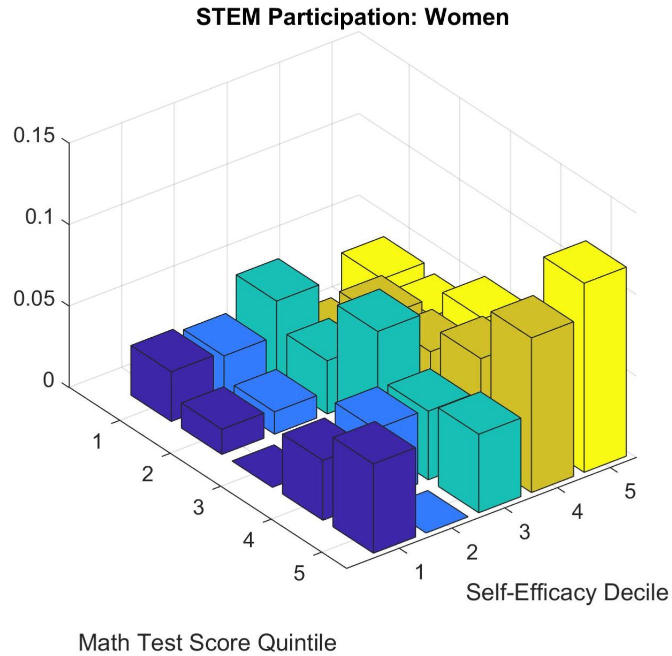
Source: Educational Longitudinal Study of 2002.
Note: Figure A.1 shows the marginal density of the math problem solving and the self-efficacy components by gender.

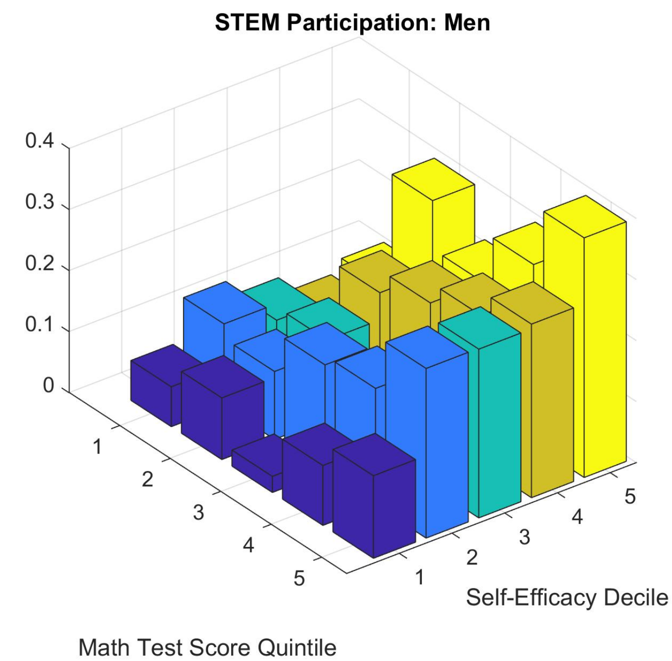
Figure A.2: Initial Major Choices by Math Ability Components for Men and Women



Source: Educational Longitudinal Study of 2002.
Note: Figure A.2 shows the marginal density of the math problem solving and the self-efficacy components by initial major choices $m_1 \in \mathcal{M}_\infty$, separately for men and for women.

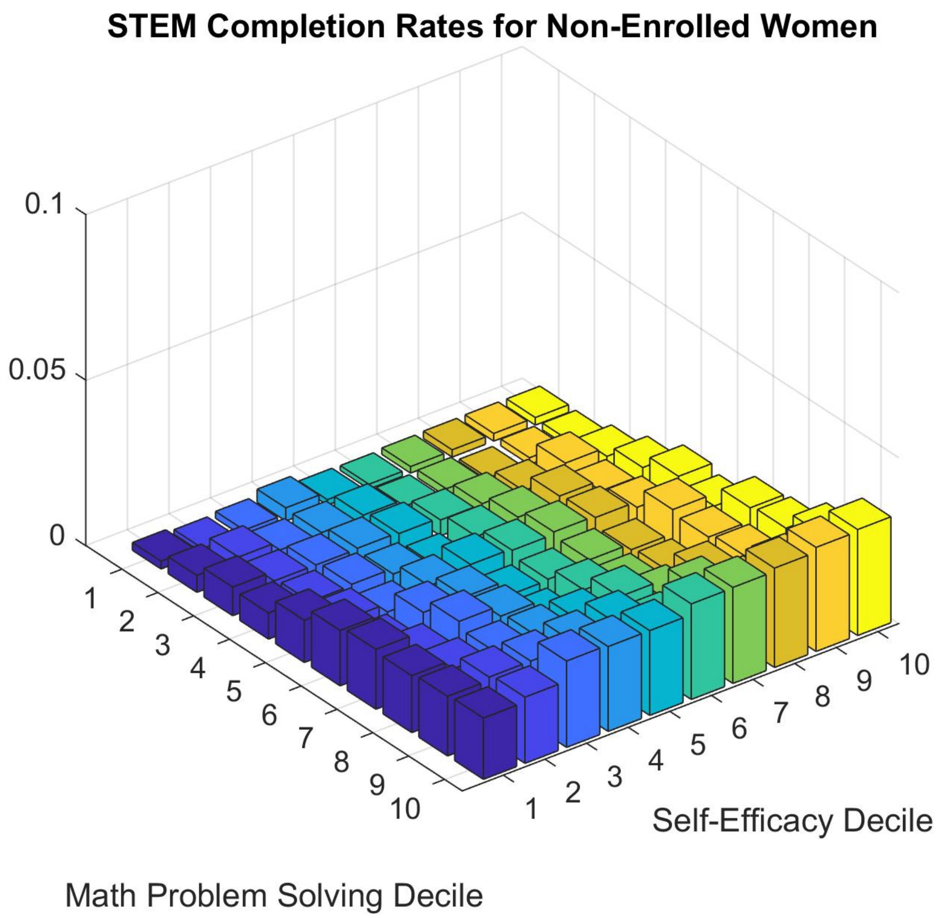
Figure A.3: Sorting into STEM Enrollment by Baseline Test Scores





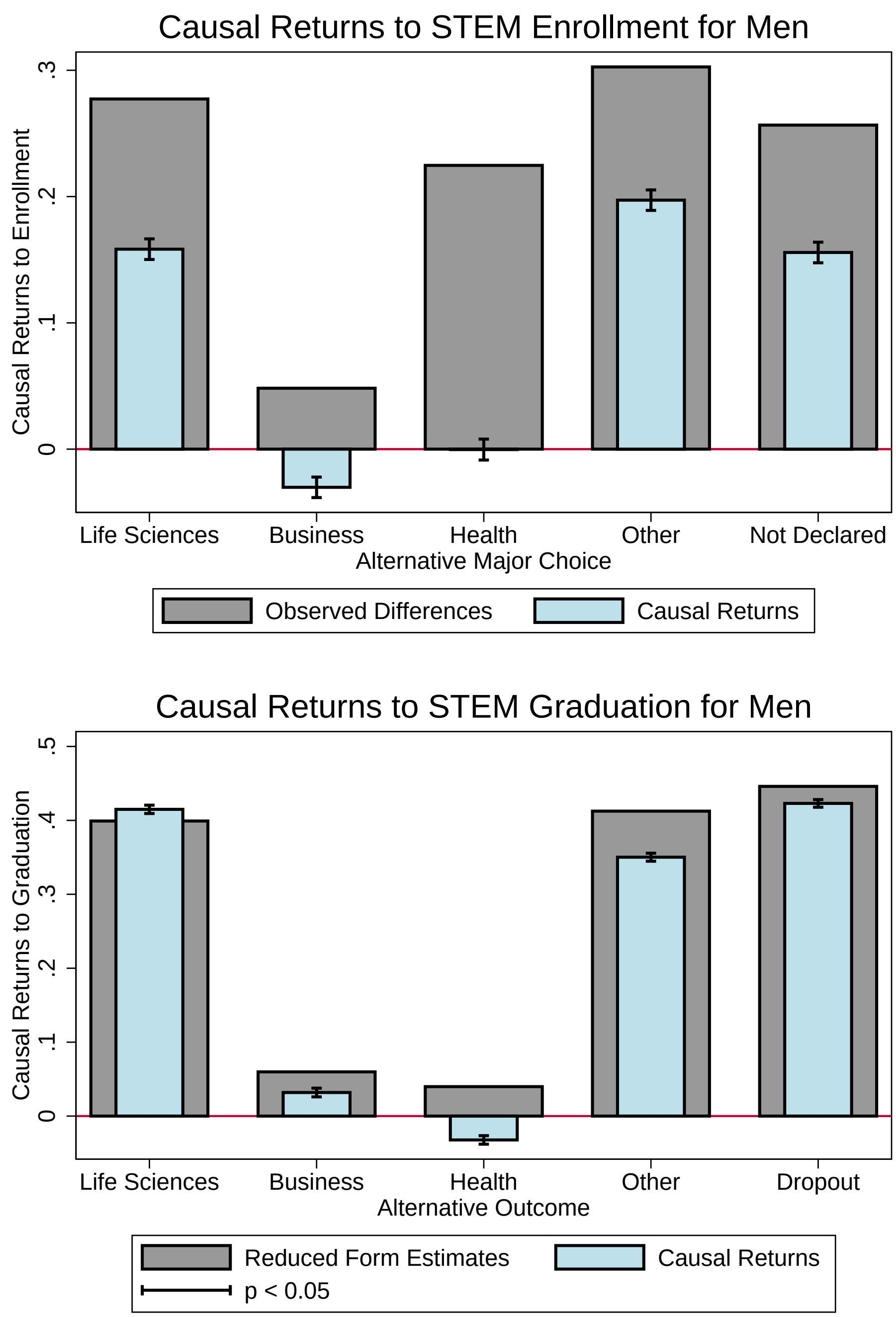
Source: Educational Longitudinal Study of 2002.
Note: Figure A.3 explores how women and sort into STEM enrollment based on baseline math test scores and self-efficacy. These measures are classified into gender-specific quintiles.

Figure A.4: STEM Completion Rates for Women Enrolled in Other Majors



Source: Educational Longitudinal Study of 2002.
Note: Figure A.4 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.

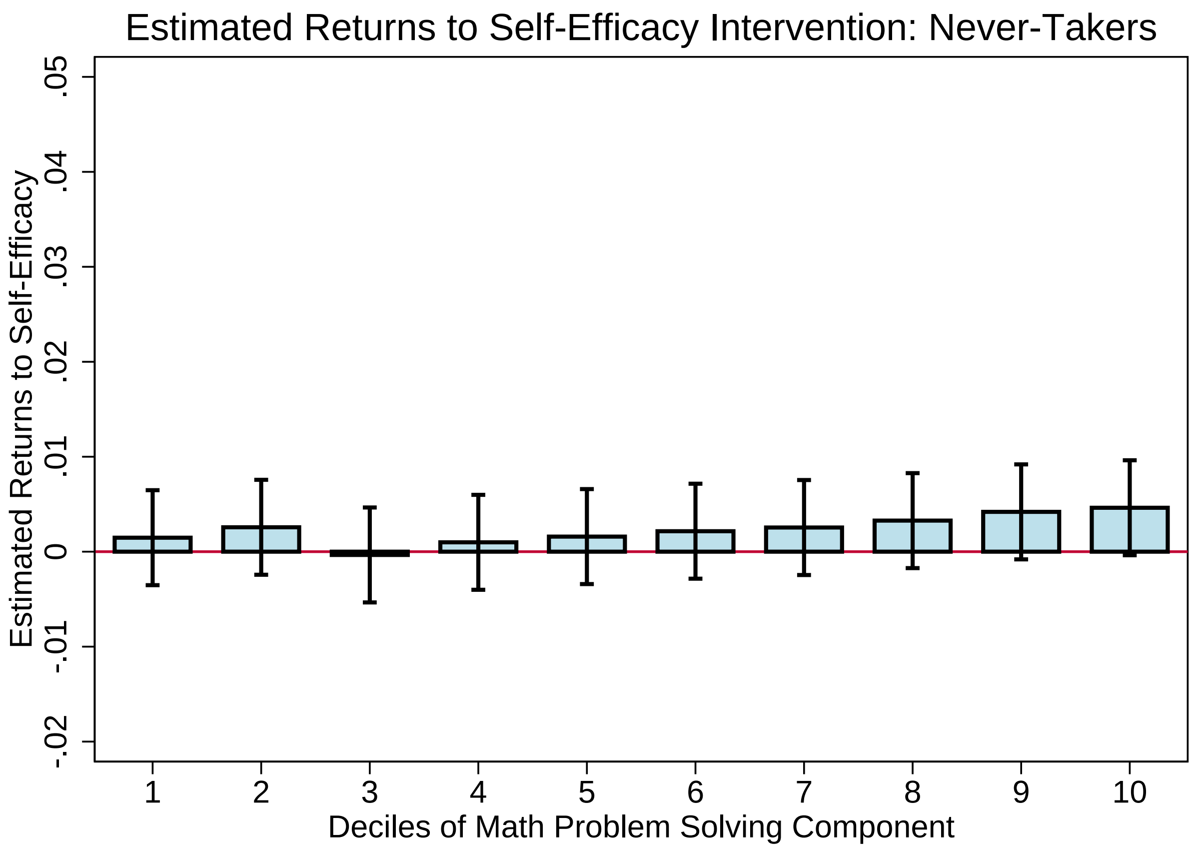
Figure A.5: Estimated Causal Returns to STEM Majors for Men



Source: Educational Longitudinal Study of 2002.

Note: Figure A.5 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 (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.

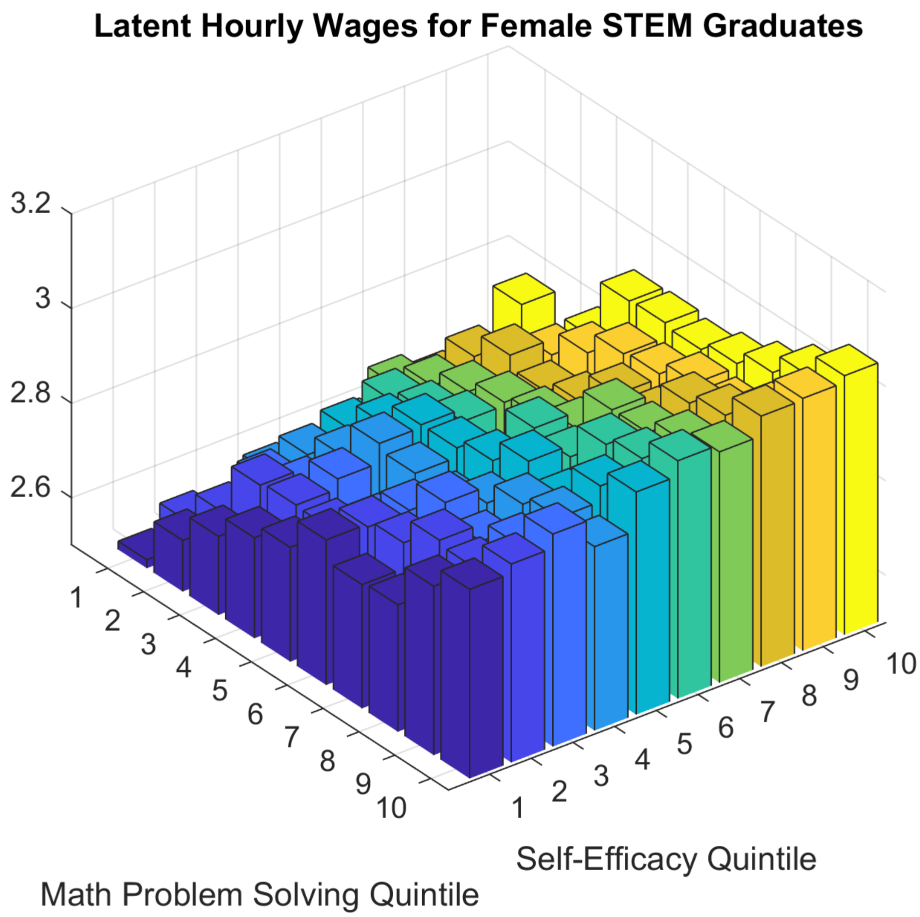
Figure A.6: Effects of Self-Efficacy Intervention on Wages for Never-Takers



Source: Educational Longitudinal Study of 2002.

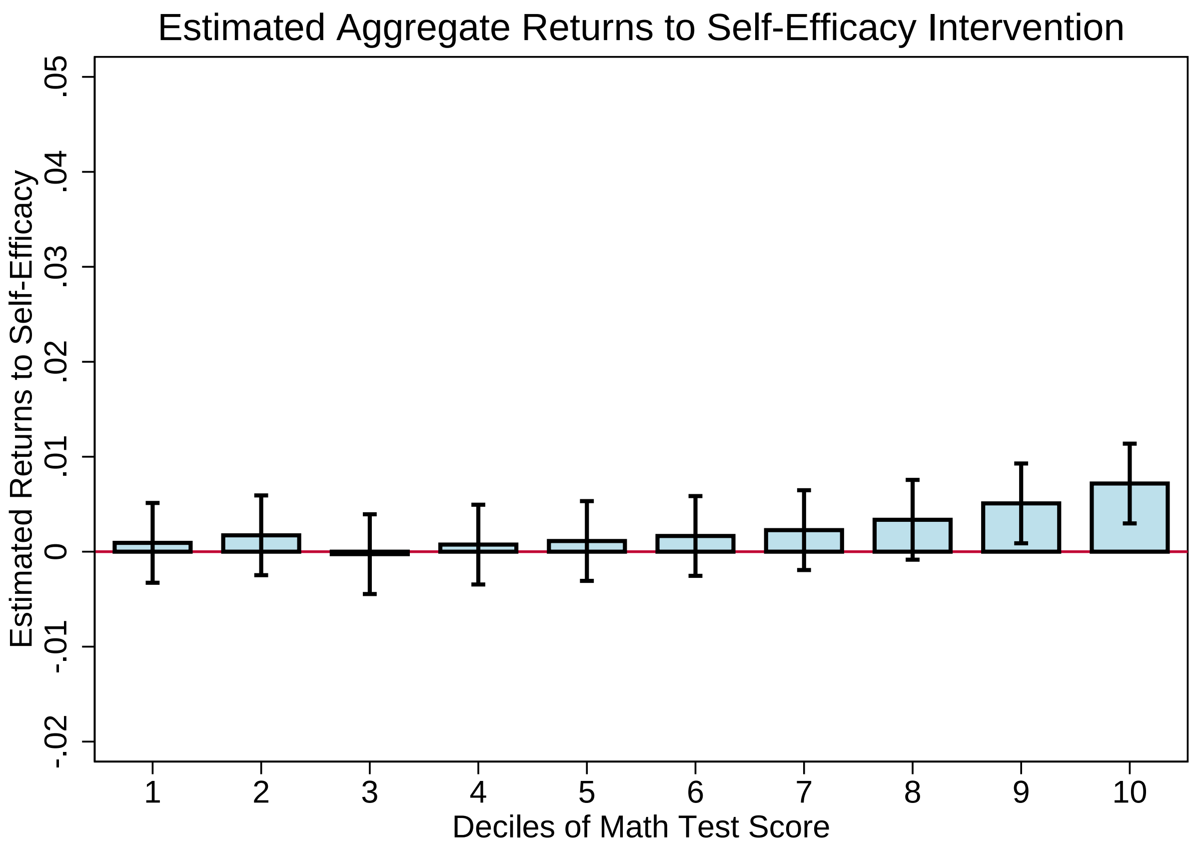
Note: Figure A.6 presents the impact of a 0.25 SD boost in θ_{SE} for the set of female college students who do not start in a STEM field and do so under the intervention, as well. I present heterogeneous impacts across the θ_C distribution. The second panel presents the same analysis for compliers, that is, the students who in the baseline *did not* enroll in STEM, but did so following the intervention.

Figure A.7: Heterogeneous Latent Wages for Female STEM Graduates



Source: Educational Longitudinal Study of 2002.
Note: Figure A.7 shows heterogeneous latent wages for women who enrolled in and completed a STEM degree by θ_C and θ_{SE} decile.

Figure A.8: Effects of Self-Efficacy Intervention on Wages by Test Score Decile



Source: Educational Longitudinal Study of 2002.
Note: Figure A.8 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.

B Identification of the Measurement System

This section presents the identification of the measurement system presented in Section 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 (7)-(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:

$$
\begin{aligned}
Cov(C_1, C_2) &= \alpha_{C_1}\alpha_{C_2}\sigma^2_{\theta,C} \\
Cov(C_2, C_3) &= \alpha_{C_2}\alpha_{C_3}\sigma^2_{\theta,C} \\
Cov(C_1, C_3) &= \alpha_{C_1}\alpha_{C_3}\sigma^2_{\theta,C}
\end{aligned}
$$

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:

$$
\begin{aligned}
Cov(C_1, G_1) &= \alpha_{C_1}\alpha_{G_1}\sigma^2_{\theta,C} \\
Cov(C_1, SE_1) &= \alpha_{C_1}\alpha_{SE_1}\sigma^2_{\theta,C} \\
Cov(C_1, R_1) &= \alpha_{C_1}\alpha_{R_1}\sigma^2_{\theta,C}
\end{aligned}
$$

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

$$
\begin{aligned}
Cov(G_1, SE_1) &= \alpha_{G_1}\alpha_{SE_1}\sigma^2_{\theta,C} + \gamma_{G_1}\gamma_{SE_1}\sigma^2_{\theta,SE} \\
Cov(G_1, SE_2) &= \alpha_{G_1}\alpha_{SE_2}\sigma^2_{\theta,C} + \gamma_{G_1}\gamma_{SE_2}\sigma^2_{\theta,SE} \\
Cov(SE_1, SE_2) &= \alpha_{SE_1}\alpha_{SE_2}\sigma^2_{\theta,C} + \gamma_{SE_1}\gamma_{SE_2}\sigma^2_{\theta,SE}
\end{aligned}
$$

$\sigma^2_{\theta,SE}$ 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 α 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:

$$
\begin{aligned}
Cov(R_1, R_2) &= \alpha_{R_1}\alpha_{R_2}\sigma^2_{\theta,C} + \eta_{R_1}\eta_{R_2}\sigma^2_{\theta,R} \\
Cov(R_1, R_3) &= \alpha_{R_1}\alpha_{R_3}\sigma^2_{\theta,C} + \eta_{R_1}\eta_{R_3}\sigma^2_{\theta,R} \\
Cov(R_2, R_3) &= \alpha_{R_2}\alpha_{R_3}\sigma^2_{\theta,C} + \eta_{R_2}\eta_{R_3}\sigma^2_{\theta,R}
\end{aligned}
$$

where $\sigma^2_{\theta,R}$ represents the variance of the self-efficacy factor. As the α_{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 (η_{R_2} and η_{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:[56]

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

I can apply Kotlarski (1967) theorem to equation (20) to non-parametrically identify the distributions of:

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

Applying the same argument to equations (7)-(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 θ_C and θ_{SE}. The correlation between both components is generated through the following linear association:

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

where θ_A is an auxiliary factor, independent of θ_C ($\theta_C \perp \theta_A$). For math GPA and self-efficacy

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

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 β_{G_1} and γ_{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 α_{G_1}, α_{SE_1}, α_{SE_2}, and α_1, which denotes the correlation between θ_C and θ_{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.[57] 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 θ_C and θ_R. I assume that the problem solving factor affects the English GPA measure only indirectly, through its correlation with the reading factor, as this measure has the lowest loading on the problem solving factor. This assumption thus allows me to identify β_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 and dropout/final major decisions, the ability loadings (α) in equations (2) and (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

[57] 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.

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 (2) and (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.

C Identification of Gender Differences in Latent Ability

In Section 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 μ_C^h and μ_C^f denote the means of the distribution of latent math problem solving ability for males and females, respectively, and Δ_C represent the difference across genders, given by $\Delta_C = \mu_C^m - \mu_C^f$.[58] Assuming that $\varphi_j^m = \varphi_j^f$, equation (25) can then be re-written as:

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

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,

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

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), 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).[59] Given this set-up, I follow Urzua (2008) and identify gender differences in Δ_C from the average across all observed math test scores and grades affected by θ_C.[60]

With Δ_C on hand, note that equation (27) has one unknown, Δ_{SE}. As a result, Δ_{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 θ_{SE} explained by each of the two observed self-efficacy measures. The same logic applies to the identification of Δ_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 Δ_C.

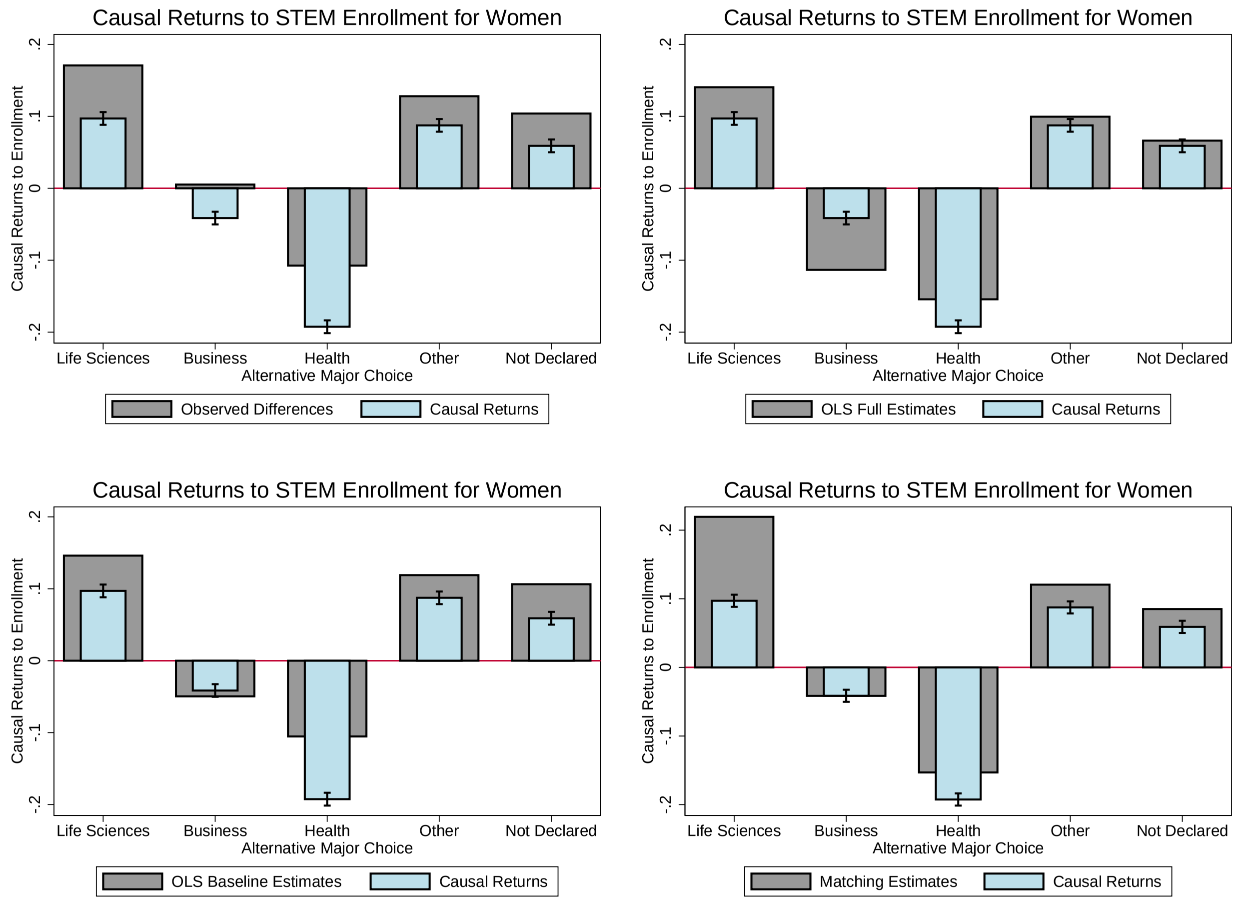
D Reduced Form Returns to College Majors

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 (13) to OLS regression and nearest-neighbor matching estimates.

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

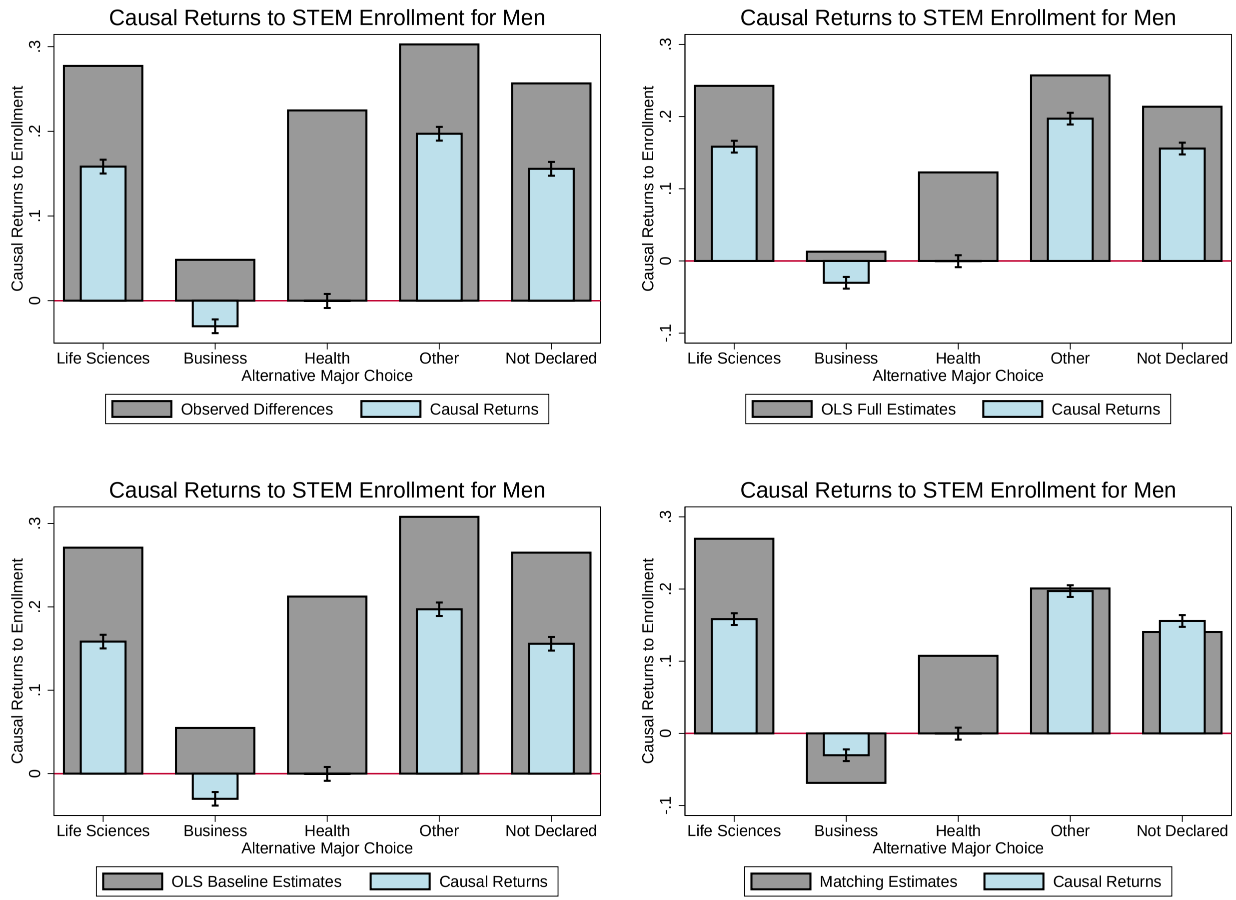
[60]Empirically, the average is calculated as a weighted average, where the weights are given by the relative share of the variance in test score measure Cj or GPA G_1 explained by θ_C. This procedure relaxes the linear average in gender differences in test scores in Urzua (2008).

Figure D.1: Reduced-Form Returns to STEM Enrollment for Women



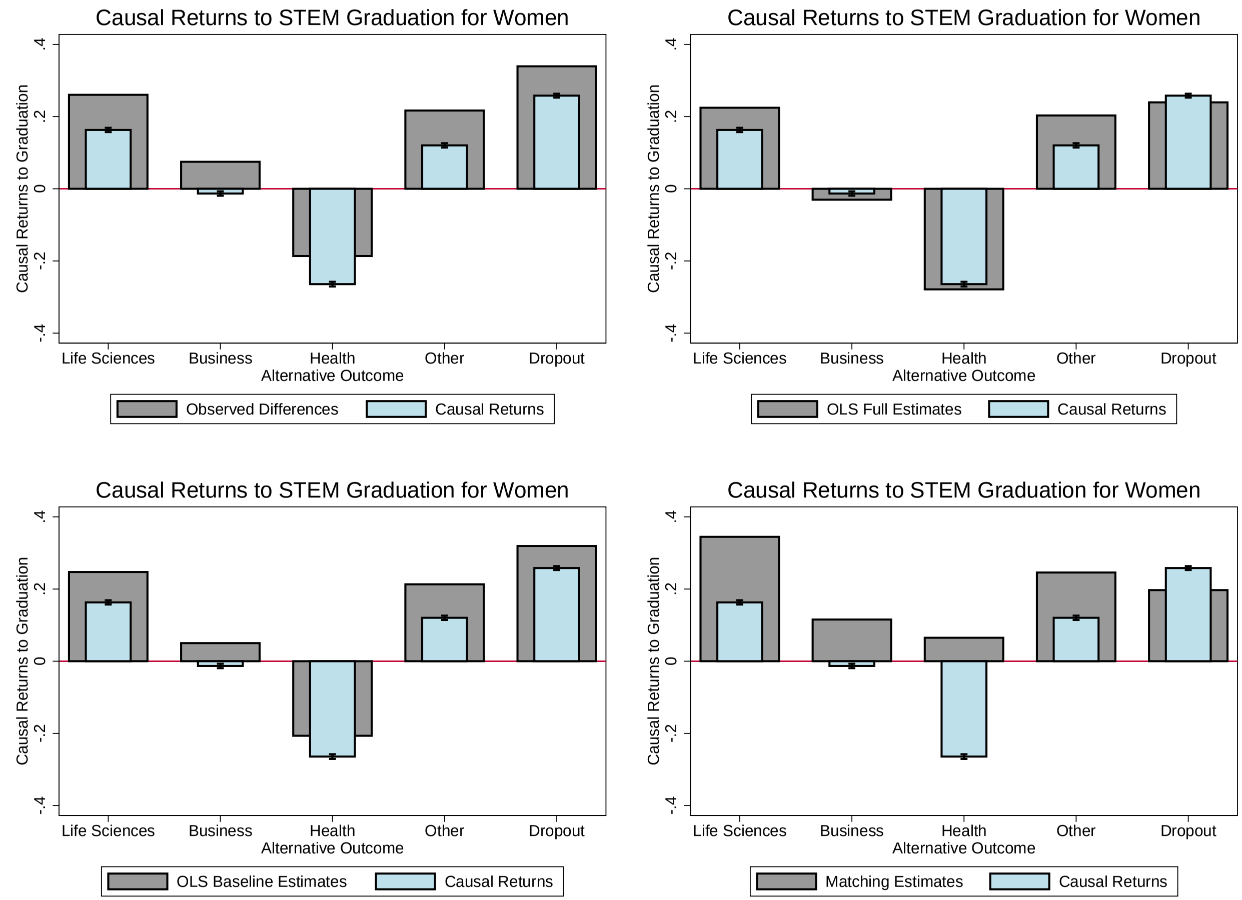
Source: Educational Longitudinal Study of 2002.
Note: Figure D.1 shows the comparison of the estimated returns to STEM enrollment against reduced form estimates. The first panel analyzes observed differences across major alternatives. The second panel estimates an OLS regression with individual and family 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 D.2: Reduced-Form Returns to STEM Enrollment for Men



Source: Educational Longitudinal Study of 2002.
Note: Figure D.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 D.3: Reduced-Form Returns to STEM Graduation for Women



Source: Educational Longitudinal Study of 2002.
Note: Figure D.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.

E "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 (1) and (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:

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

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). ε is an arbitrarily small neighborhood around the margin of indifference.[61] As the 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.

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

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.[62] 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.

[62] 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 E.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
θ_C	0.325	0.395	0.366	0.010	0.332	0.268
θ_{SE}	0.417	0.601	0.352	0.086	0.398	0.391
θ_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

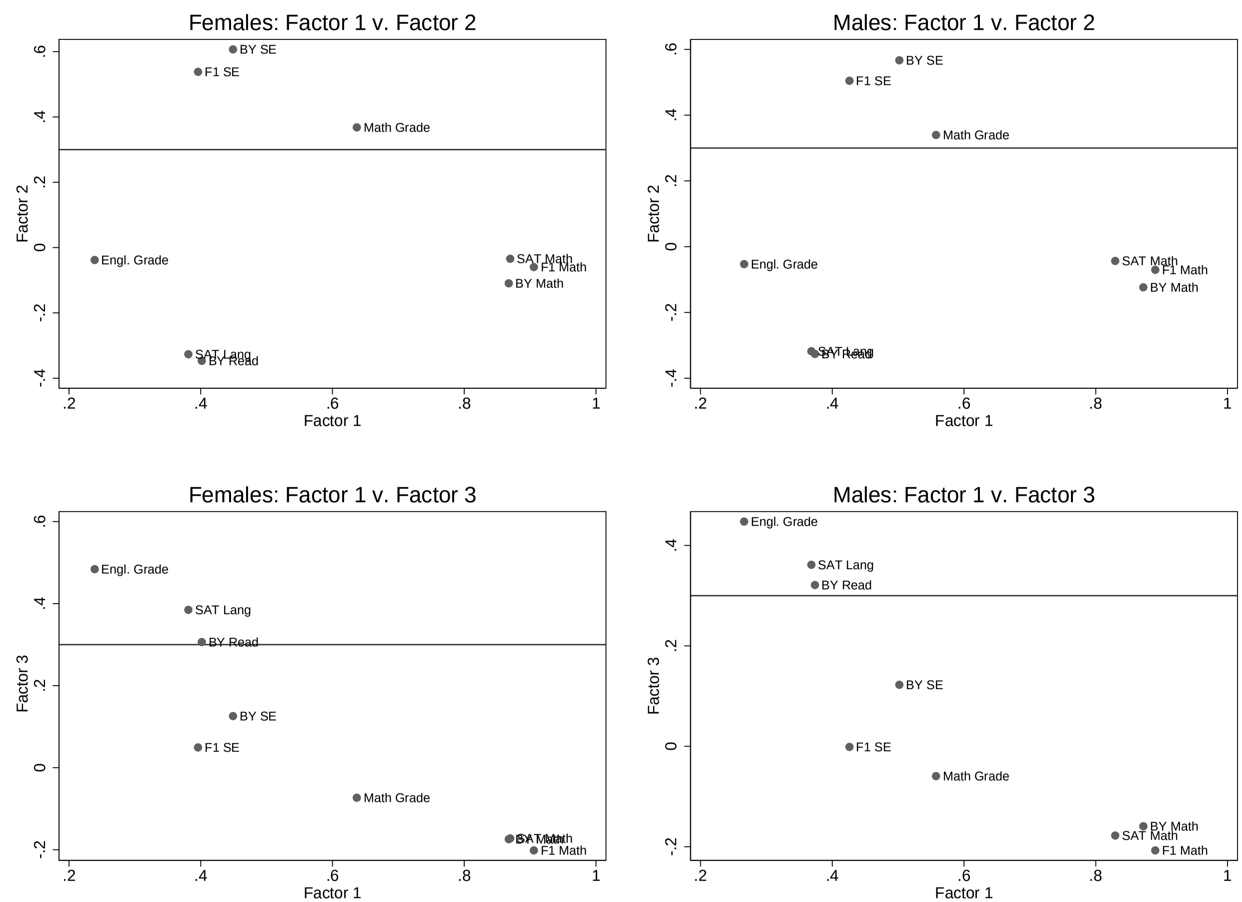
Panel B. Final Majors among STEM Enrollees

	$ID_{m_1,G}$	$ID_{m_1,G,S}$	$ID_{m_1,G,D}$
	Full Set	Major Switchers	Dropout
	(1)	(2)	(3)
Both Parents	0.845	0.839	0.855
Family Income	11.30	11.32	11.26
θ_C	0.362	0.409	0.286
θ_{SE}	0.440	0.560	0.247
θ_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. θ_C represents the problem solving factor, θ_{SE} is the math self-efficacy component and θ_R is the reading ability component. Results are simulated from the estimates of the model.

F Exploratory Factor Analysis

Figure F.1: Loadings from Factor Analysis



Source: Educational Longitudinal Study of 2002. Note: Figure F.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.