Multidimensional Skills and Gender Differences in STEM Majors

Fernando Saltiel*

McGill University

August 15, 2022

Abstract

This paper studies the relationship between pre-college skills and gender differences in STEM majors. I use longitudinal data to estimate a generalized Roy model of initial major choices and subsequent graduation outcomes. I recover students' latent math ability, non-cognitive skills and math self-efficacy. High math ability women have lower math self-efficacy than men. Mathematical ability and self-efficacy shape the likelihood of STEM enrollment. A lack of math self-efficacy drives women's drop out from STEM majors. I find large returns to STEM enrollment for high math-ability women. Well-focused math self-efficacy interventions could improve women's STEM graduation rates and labor market outcomes.

Keywords: Generalized Roy Model, Major Choices, Non-Cognitive Skills, Gender Gaps.

JEL Codes: J01, J24, I24, I26, J16.

*Email: fernando.saltiel@mcgill.ca. Department of Economics, McGill University, Leacock Building 512, 855 Sherbrooke St. West, Montreal, QC H3A 2T7. I am grateful to the Editor (Barbara Petrongolo) and two anonymous referees for helpful comments and suggestions. Sergio Urzúa provided invaluable guidance and support in this project. Melissa Kearney, Judy Hellerstein and Nolan Pope contributed important insights as part of my dissertation committee. I thank numerous seminar and conference participants for helpful comments and conversations. An earlier version of this paper was the first chapter of my dissertation, previously entitled "What's Math Got to Do With It? Multidimensional Ability and the Gender Gap in STEM". Parts of this paper were written as a post-doctoral associate at Duke University. All errors are my own.

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 (Webber, 2014; Patnaik et al., 2020), examining the drivers of women's participation in STEM majors is an important step towards improving their labor market outcomes. In this context, colleges across the country have begun implementing policies aimed at boosting women's STEM enrollment rates (Olson and Riordan, 2012). 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 (Astorne-Figari and Speer, 2019). As a result, understanding the factors which shape students' initial and final major choices can help in designing more effective policies to promote STEM participation and persistence.

In this paper, I examine the interaction between multidimensional skills and college major choices, focusing on women's enrollment and graduation from math-intensive STEM majors. The existing literature has examined the how students' test scores affect STEM participation rates (Turner and Bowen, 1999; Speer, 2017) and extensively analyzed the importance of preferences and beliefs in shaping major choices (Arcidiacono, 2004; Zafar, 2013; Wiswall and Zafar, 2015, 2018; Reuben et al., 2017; Patnaik et al., 2020). However, test scores are affected by background characteristics and contaminated with measurement error (Borghans et al., 2008; Kautz et al., 2014), thus potentially mismeasuring the importance of skills in shaping STEM participation. Moreover, other skill dimensions may play an important role in determining students' major choices.

In this context, I consider mathematical self-efficacy as a potential driver of gender differences in STEM participation. Math self-efficacy measures a student's perception of her capacity to succeed at math-related problems and courses, and efficacious students are more likely to pursue challenging math courses, exert greater effort in class and persist in solving difficult math problems (Betz and Hackett, 1983).[1] As such, having higher math self-efficacy may increase the likelihood that students pursue STEM majors and that they successfully persist through graduation.

To understand how multidimensional skills shape students' educational attainment, I present and estimate a generalized Roy model, which encompasses sequential decisions of college major choices and subsequent completion outcomes. In this model, which builds on Heckman et al. (2016, 2018), students first select a college major among five broad fields. Students subsequently decide whether to remain in college or to drop out, and continuers lastly decide whether to complete their initial degree or switch majors. Upon completing their studies, students enter the labor market, decide whether to work, and if they choose to do so, earn hourly wages. Throughout the model, individual decisions and labor market outcomes are a function of observed characteristics and students' multidimensional skills, encompassing their math ability, non-cognitive skills and mathematical self-efficacy. I thus distinguish non-cognitive skills capturing students' grit and persistence from

[1] The concept of self-efficacy was introduced by Bandura (1977) and it is defined as "beliefs in one's capabilities to organize and execute the courses of action required to manage prospective situations". A person's self-efficacy is domain-specific and it measures a person's perception of her capacity to succeed at specific tasks in that domain.

their self-efficacy in mathematics. I estimate the model separately for men and women to capture gender differences in the interaction between skills, major choices and labor market outcomes.

I implement the model using Educational Longitudinal Study of 2002 (ELS) data, which follows a nationally-representative cohort of 10^{th} graders for ten years. ELS data includes detailed information on multiple measures of math test scores, high school grades, non-cognitive skills measures — encompassing students' action control, control expectation and their instrumental motivation, multiple math self-efficacy measures, detailed information on college major choices and early-career labor market outcomes, capturing respondents' hourly wages at age 25.

I posit a measurement system of observed skill measures to recover the distribution of multi-dimensional latent abilities, following Carneiro et al. (2003); Hansen et al. (2004); Heckman et al. (2006), among others. This approach allows me to correct for measurement error in test scores while controlling for the contribution of background characteristics to observed skill measures. I take advantage of the various observed measures in the data to recover students' latent math ability, a non-cognitive skill factor and their latent mathematical self-efficacy, allowing for the latent factors to be correlated (Prada and Urzúa, 2017). First, I find that men have higher math self-efficacy than women, in the range of 0.4 standard deviations, and no discernible gender gaps emerge in the other two skill dimensions. Importantly, while math ability is positively correlated with self-efficacy, the correlation is far higher for men (0.488) than for women (0.304). Altogether, these results indicate that among high-math-skilled students, a far lower share of female students have high self-efficacy in math compared to their male peers.

Both math ability and self-efficacy are strong predictors of STEM enrollment for both men and women. For instance, for women in the top math ability decile, only 5.5% of those who are in the bottom self-efficacy decile enroll in STEM, whereas 13.3% of those in the top decile do. As a result, the relative lack of women at the top of the joint skill distribution reduces their participation in math-intensive majors. I similarly find that among the highest math-ability males, self-efficacy also increases their STEM participation, as 35.4% of men in the top joint decile of math ability and self-efficacy enroll in STEM. As such, a sizable gender gap in STEM participation remains conditional on latent skills, fitting in with prior work highlighting the importance of preferences in driving STEM gaps (Zafar, 2013; Wiswall and Zafar, 2015, 2018). On the other hand, an increase in women's mathematical self-efficacy by a full standard deviation would increase their STEM enrollment rates by upwards of 45%, thus highlighting the importance of considering multiple skill dimensions when analyzing college major choices.

In terms of subsequent STEM completion, while 62% of men initially enrolled in these majors end up graduating, just 55% of women do so. There is re-sorting on math ability for both men and women, such that only the highest math-skilled students graduate from these majors. However, self-efficacy plays a far larger role for women than it does for men in leading to degree completion. 44% of male enrollees in the bottom self-efficacy quintile complete a STEM degree, rising to 67% for those in the top quintile. On the other hand, while only 29% of female STEM enrollees in the bottom self-efficacy quintile successfully complete a degree, the completion rate for those in the top quintile more than doubles, exceeding 66%. Moreover, I do not find evidence that higher non-cognitive skills increase the likelihood of STEM completion for women, thus remarking the importance of extending the analysis to incorporate additional dimensions of students' skills besides those traditionally incorporated in the literature on non-cognitive skills (Heckman et al., 2006). All in all, a shortfall in math self-efficacy reduces the likelihood that women enroll and complete a STEM major, thus contributing to gender differences in STEM participation rates.

Despite the efforts aimed at increasing women's STEM participation, the extent to which all female students would enjoy positive returns from pursuing these majors remains an open question (Altonji et al., 2012, 2016). In this context, the model allows me to recover potential wages across initial majors, and since wage outcomes also depend on students' latent abilities, I can estimate gender-specific heterogeneous returns to STEM enrollment across the ability distribution. I show that 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, the average returns against business and health fields are negative.[2] On the other hand, I find significant heterogeneity in these returns, such that high math ability women would largely benefit from enrolling in STEM. Moreover, as I approximate the latent utilities associated with each initial major, I can identify students' second-best majors (Kirkeboen et al., 2016). High math-ability women in other majors with a next-best option in STEM would have enjoyed positive returns to STEM enrollment instead. I lastly estimate the conditional returns to STEM graduation after enrollment, finding all women would benefit from finishing these degrees relative to switching to other majors or to dropping out from college. The returns to conditional STEM completion are strongly increasing in the math ability distribution.

Lastly, the importance of math self-efficacy in predicting women's STEM participation, coupled with the malleability of non-cognitive skills through adolescence (Kautz et al., 2014), suggests that interventions aimed at boosting self-efficacy could have a sizable impact on women's STEM participation rates. Using the estimated model parameters, I show that increasing high-math-ability women's self-efficacy by 0.25 standard deviations would lead to increased STEM enrollment and graduation by 10-15 percent relative to baseline participation rates. In light of the positive returns to STEM enrollment and completion for high-skilled women, I find that the simulated math self-efficacy increase would lead to small increases in women's wages in the early career.

This paper contributes to various strands of the literature, standing at the intersection of prior work analyzing the importance of non-cognitive skills, college major choices and gender differences in educational attainment. First, a number of important papers have examined the drivers of students' college major choices, including the importance of pre-college skills in shaping initial major choices and subsequent completion outcomes (Altonji, 1993; Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2014; Kinsler and Pavan, 2015; Arcidiacono et al., 2016; Humphries et al., 2017, 2019). I contribute to this literature by analyzing how multiple dimensions of students' non-

[2]The returns to major are estimated using evidence on age-25 hourly wages. These outcomes do not capture the full extent of lifecycle returns to college majors (Webber, 2014; Altonji et al., 2016; Patnaik et al., 2020).

cognitive skills affect their major choices and conditional completion outcomes, while documenting heterogeneous impacts by gender. This paper further fits in with an extensive literature examining the drivers of gender differences in major choices, including the importance of high school preparation (Joensen and Nielsen, 2016; Shi, 2018; Aucejo and James, 2021; Card and Payne, 2021), test scores and cognitive skills (Turner and Bowen, 1999; Dickson, 2010; Speer, 2017; Astorne-Figari and Speer, 2019; Jiang, 2021), preferences and beliefs (Zafar, 2013; Wiswall and Zafar, 2015, 2018; Delaney and Devereux, 2019; Bordón et al., 2020; Kugler et al., 2021; Ahn et al., 2022), teachers (Carrell et al., 2010; Carlana, 2019) and peers (Fischer, 2017; Mouganie and Wang, 2020; Brenøe and Zölitz, 2020). I extend this literature by demonstrating the importance mathematical self-efficacy in shaping differential STEM enrollment by gender and by showing its importance for STEM completion rates among women.

I further contribute to an extensive literature which has estimated the returns to majors using a variety of approaches. These identification strategies encompass linear regressions (Rumberger and Thomas, 1993; Chevalier, 2011; Webber, 2014; Deming and Noray, 2020), varied structural approaches (Arcidiacono, 2004; Kinsler and Pavan, 2015; Humphries et al., 2019; Mourifie et al., 2020) and regression discontinuity designs (Hastings et al., 2013; Kirkeboen et al., 2016). I present novel evidence by recovering gender-specific returns to enrolling in STEM majors, estimating heterogeneous returns against second-best majors and presenting conditional returns to STEM completion.

Lastly, this paper further contributes to a growing literature on the importance of non-cognitive skills in shaping educational and labor market outcomes (Heckman et al., 2006; Lindqvist and Vestman, 2011). Recent work has identified the importance of different dimensions of the non-cognitive skill vector: Humphries et al. (2019) separately identify grit and interpersonal skills and Humphries and Kosse (2017) distinguish non-cognitive skills from preferences and personality traits. I add to this literature by showing the differential importance of math self-efficacy in shaping major choices relative to the estimated importance of 'traditional' non-cognitive skill constructs. I further show the importance of this distinction for understanding gender gaps in major choices.

The rest of the paper is structured as follows. In Section 2, I describe the data sources and present descriptive evidence on the drivers of college major choices. In Section 3, I introduce the model of college major choices, along with the estimation approach. In Section 4, I present evidence on the latent factors and on sorting patterns into initial majors and final educational outcomes. In Section 5, I present the gender-specific returns to majors and the conditional returns to completing such majors. In Section 6, I present the estimated impacts from a simulated intervention aimed at boosting women's math self-efficacy. I conclude and discuss my results in Section 7.

2 Data Sources and Summary Statistics

2.1 Data Sources

This paper uses longitudinal data from the Educational Longitudinal Survey (ELS) of 2002 (Ingels et al., 2014). 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. The first two surveys include detailed information on students' individual characteristics, including their race and gender, family characteristics, including family composition, parents' educational attainment and total family income. Moreover, ELS data includes multiple measures of students' academic performance, including their high school GPA, their performance on 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 additionally includes various questions in the baseline survey measuring respondents' non-cognitive skills. These questions capture students' expectations of success in academic learning (control expectation scale), their motivation to perform well academically in order to reach external goals like future job opportunities or financial security (instrumental motivation) and their perceived effort and persistence when facing difficulties (action control), which is closely related to grit (Duckworth et al., 2007).[3]

Self-Efficacy. As discussed in the introduction, self-efficacy captures a person's perceived capacity to accomplish a specific task.[4] Assessments of self-efficacy must be 'domain-specific' in order to correctly measure this concept (Pajares, 1996; Bandura, 1997). The baseline survey of the ELS thus included a battery of questions regarding students' self-efficacy in Mathematics and English. In particular, students in the baseline survey were prompted to answer the following questions:

1. Confident I can do an excellent job on my Math/English tests.
2. Certain I can understand the most difficult material presented in Math/English texts.
3. Confident I can understand the most complex material presented by my Math/English teacher.
4. Confident I can do an excellent job on my Math/English assignments.
5. Certain I can master the skills being taught in my Math/English class.

These questions were answered on a four-point Likert scale encompassing almost never, sometimes, often, and almost always as options. Importantly, men are far more likely than women to respond they are 'almost always' confident in their math-related tasks than women, yet this is not the case in English (Table A.1). In light of the gender gaps in math self-efficacy and given my interest in understanding the drivers of STEM participation, I focus on mathematical self-efficacy for the rest of the paper. I additionally rely on students' responses to math self-efficacy questions in the first follow-up survey, and create a math self-efficacy variable in each survey round after applying principal component analysis to the five underlying questions.[5]

[3]These three 'traditional' non-cognitive skill measures are constructed by ELS staff using exploratory factor analysis on the responses to specific questions (Saltiel, 2020). For the descriptive analysis presented below, I construct a non-cognitive skills index from a principal components analysis (PCA) across these three measures.

[4]Mastery experience — where students interpret their past performance to revise their judgments of competence in that task — is the main driver of their self-efficacy (Bandura, 1997).

[5]To examine whether the math self-efficacy and non-cognitive variables capture different underlying constructs, I perform an exploratory factor analysis (EFA) using the five measures in the ELS (Figure A.1). Assuming orthogonal

Sample Selection. Since the goal of this paper is to understand the interaction between skills and college major choices, I restrict my sample to students enrolled in four-year college by the second follow-up survey (age 20).[6] Nonetheless, the final sample includes students who do not graduate with a four-year degree, bachelor's recipients and students in post-graduate education. I consider students' progression through college majors by first using their reported major in the second follow-up survey, including those who had not yet declared one. For students who had earned a Bachelor's degree by 2012, I examine their final major at graduation using information from their college transcripts. 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. Since the STEM gender gap is largely driven by differences in math-intensive fields (Kahn and Ginther, 2017), I group majors into five categories, which include math-intensive STEM, life sciences, business, health, and the remaining majors.[7]

I analyze respondents' labor market outcomes using information reported in the third follow-up survey. Students report information on their labor market outcomes in 2011, covering age-25 outcomes for the majority of the sample. Respondents indicate whether they worked, the number of weeks and hours per week they were employed and their total employment earnings during the year. I use these variables to construct a measure of hourly wages for each individual in the sample.[8]

I first restrict the baseline ELS sample to include students who take the baseline exams along with those who report at least one valid math self-efficacy measure.[9] Restricting the sample to four-year college enrollees substantially reduces the sample to 2,899 women and 2,284 men, fitting in with higher rates of college enrollment for women (Goldin et al., 2006). I lastly drop individuals who do not report their final educational attainment or valid labor market outcomes in the endline survey, yielding a final sample of 4,599 students, encompassing 2,615 women and 1,984 men.[10]

factors, EFA indicates the existence of two factors. The first factor loads on the three 'traditional' non-cognitive skill measures. The second encompasses both math self-efficacy measures, action control and control expectation. Instrumental motivation only loads on the first factor. These results inform the measurement system in Section 3.2.

[6] I examine whether men and women are differentially selected into the four-year college enrollee sample on their baseline test scores (Table B.1). Higher-skilled students are more likely to initially enroll in four-year college, but such sorting patterns largely do not vary by gender.

[7] Math-intensive STEM fields include degrees in engineering, computer science, mathematics, statistics and physics. Life science majors encompass biology degrees. Business degrees include majors in business, management and marketing. Health includes majors in clinical sciences and for health professionals. The "Other" group includes the remaining majors, and individuals who had not yet declared an initial major.

[8] To avoid including part-time workers in the returns to majors, I examine wage outcomes for individuals who reported having worked at least 500 hours in 2011 and earned at least five dollars an hour. Wages are top coded at $125/hour. Results are robust to alternative restrictions. The framework in Section 3 models employment decisions.

[9] I observe the full set of observed skill measures for 60.4% of the sample. Observed characteristics are only weakly predictive of the likelihood of having an observed measure and insignificant across a number of measures (Table A.2). In the reduced form analysis, I impute the sample average for individuals with missing test scores and include a dummy variable to account for non-response. As discussed in Section 3, the model is identified despite small differences in non-respondents' observed characteristics.

[10] Table A.3 outlines how the various sample restrictions result in the final sample used in the paper.

2.2 Reduced-Form Evidence

Table 1: Baseline Characteristics

	Full Sample (1)	STEM (2)	Life Sciences (3)	Business (4)	Health (5)	Other (6)
Panel A. Women						
Background Characteristics						
Both Parents	0.81	0.81	0.85	0.79	0.82	0.82
Parental Education	15.67	16.31	15.97	15.37***	15.36***	15.72***
HH Income	0.65	0.71	0.64	0.61*	0.61*	0.66
Underrepresented Minority	0.18	0.22	0.20	0.19	0.20	0.18
Skill Measures						
Baseline Math Exam	-0.11	0.28	0.15	-0.19***	-0.28***	-0.11***
HS GPA	0.13	0.31	0.39	0.06**	0.05**	0.11**
Baseline Math Self-Efficacy	-0.16	0.29	0.16	-0.04***	-0.15***	-0.25***
Non-Cognitive Skills (PCA)	0.02	0.21	0.33	-0.04**	0.02	-0.02**
Educational Outcomes						
College Dropout	0.15	0.18	0.13	0.18	0.18	0.14
Complete Initial Major	0.67	0.55	0.51	0.65**	0.46	0.74***
Labor Market Outcomes						
Employed	0.77	0.77	0.65**	0.81	0.77	0.77
Hourly Wages	18.69	20.19	16.92*	19.43	22.97*	17.65**
Observations	2,615	119	176	297	370	1,653
		4.6%	6.7%	11.4%	14.1%	63.2%
Panel B. Men						
	Full Sample (1)	STEM (2)	Life Sciences (3)	Business (4)	Health (5)	Other (6)
Background Characteristics						
Both Parents	0.83	0.83	0.82	0.87	0.87	0.82
Parental Education	15.91	15.99	16.50**	15.91	15.16***	15.86
Family Income (Log)	0.70	0.70	0.71	0.76	0.55**	0.69
Underrepresented Minority	0.16	0.18	0.17	0.16	0.14	0.15
Skill Measures						
Baseline Math Exam	0.15	0.47	0.45	0.06***	-0.20***	0.05***
HS GPA	-0.16	0.12	0.22	-0.23***	0.05	-0.30***
Baseline Math Self-Efficacy	0.23	0.60	0.42*	0.17***	0.19***	0.10***
Non-Cognitive Skills (PCA)	-0.02	0.08	0.18	-0.02	0.17	-0.10**
Educational Outcomes						
College Dropout	0.17	0.14	0.13	0.12	0.14	0.20**
Complete Initial Major	0.60	0.62	0.53*	0.70**	0.22***	0.60
Labor Market Outcomes						
Employed	0.77	0.81	0.59***	0.85	0.71*	0.75**
Hourly Wages	20.02	23.84	18.85***	21.06***	25.81	18.01***
Observations	1,984	369	127	335	69	1,084
		18.6%	6.4%	16.9%	3.5%	54.6%

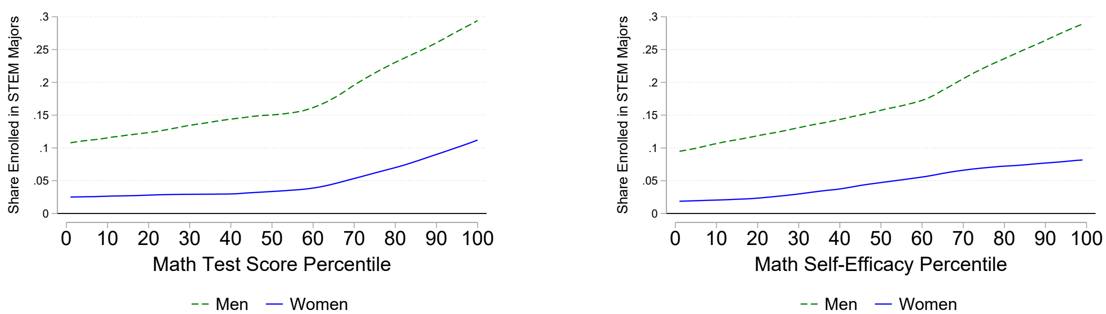
Notes: Table 1 presents summary statistics for the main sample of female and male four-year college enrollees considered in the paper. HH income is a binary variable that equals one for students in households above the U.S. household median income in 2002. All test score and non-cognitive skill measures are standardized in the full sample. Educational outcomes are observed in the endline survey round, conducted in 2012. Employed individuals are those who worked at least 500 hours in 2011 and hourly wages are calculated by dividing total employment earnings by the number of hours worked. The last five columns present averages for students across each initial major and the stars in last four columns indicate the difference of students enrolled in STEM majors relative to those in the life sciences, business, health and other majors, respectively, following from a two-sided t-test. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Major Choices. In the first and second panels of Table 1, I present summary statistics for the women and men included in the main sample. The first column presents average characteristics by gender. Skill measures, standardized in the full sample, show men outpace women in math exam performance, whereas women earn higher grades in high school (Goldin et al., 2006; Pope and Sydnor, 2010). Despite no discernible gender gaps in non-cognitive skills, I find significant differences in math self-efficacy, as men's self-efficacy exceeds women's by 0.39 standard deviations.

The last row in each panel documents the number of students enrolled across initial majors, demonstrating sizable gender differences in the prevalence of STEM enrollment, as just 4.6% of women start in these majors, compared to 18.6% of men. Both men and women enrolled in STEM tend to come from higher-educated households vis-à-vis their peers in other majors. Importantly, STEM enrollees outpace their counterparts in other majors in the baseline math test score as well as in math self-efficacy, with students in life sciences earning the second-highest scores in these two measures. In fact, both men and women exhibit stronger sorting-into-STEM on their math self-efficacy than on their non-cognitive skills, remarking the importance of analyzing different dimensions of students' skills as a driver of major choices.[11] Lastly, a larger proportion of women complete their initial major than men on average, yet this pattern is reversed in math-intensive STEM fields, as just 55% of female STEM enrollees complete their majors, compared to 62% of their male counterparts.

Figure 1 further shows that male and female students at the top of the math test score distribution are far more likely to enroll in STEM, yet large gender differences remain across the entire distribution. The second panel shows that similar patterns remain across the math self-efficacy distribution.

Figure 1: STEM Enrollment Rates by Math Test Scores and Self-Efficacy



(a) Math Test Score **(b)** Math Self-Efficacy

Note: Figure 1 presents the share of STEM enrollees by gender by their percentile in the math test scores (Panel A) and mathematical self-efficacy (Panel B) distributions. The math test score and self-efficacy measures are each constructed through principal component analysis using the respective baseline and follow-up variables.

Labor Market Outcomes. Table 1 shows that health fields are the highest-paid in the early career for both genders, fitting in with Altonji et al. (2012). At the same time, STEM enrollees earn higher average wages compared to their peers in the remaining majors, and these differences are significant with respect to students in life sciences and 'Other' degrees.

[11] Math test scores and self-efficacy predict STEM enrollment in a multinomial logit that directly controls for other skill measures and background characteristics (Table A.4). These patterns are robust to controlling for reading test scores (Table A.5), as Aucejo and James (2021) had found that women's relative advantage in verbal skills contributes to the STEM gender gap in the UK.

Table 2: OLS Estimates: Wage Returns to Majors by Gender

	Women	Women	Men	Men
	(1)	(2)	(3)	(4)
STEM	0.060	0.021	0.270***	0.250***
	(0.063)	(0.062)	(0.034)	(0.035)
Life Sciences	-0.042	-0.068	0.005	-0.014
	(0.047)	(0.047)	(0.065)	(0.065)
Business	0.107***	0.104***	0.173***	0.167***
	(0.031)	(0.031)	(0.032)	(0.032)
Health	0.253***	0.257***	0.323***	0.307***
	(0.033)	(0.033)	(0.083)	(0.083)
Background Characteristics	Yes	Yes	Yes	Yes
Test Scores	No	Yes	No	Yes
Observations	2,007	2,007	1,520	1,520

Note: Table 2 presents evidence from a linear wage regression examining the returns to enrolling in different majors for women (first two columns) and for men (last two columns). The first and third columns include individual and family background characteristics as control variables. The second and fourth columns add baseline test scores as controls. The omitted category are students in 'Other' majors. The sample includes individuals in the estimation sample who worked at least 500 hours in 2011 and earned at least five dollars per hour. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

However, as students sort into STEM based on their observed characteristics and skills, these wage differences do not capture the returns to such majors. In Table 2, I thus estimate an OLS regression to explore the returns to majors upon controlling for background characteristics and baseline skills. Female health enrollees enjoy a larger wage premia than those in STEM even when controlling for baseline skill measures (column 2), whereas male STEM enrollees outearn their peers in life sciences, business and 'Other majors' in the corresponding specification (column 4).

These results represent a first approximation towards understanding the gender-specific returns to college majors, yet they rely on a strong selection-on-observables assumption. To recover the returns to college majors, I introduce a discrete choice model which accounts for endogenous selection of college majors, final educational attainment and labor market outcomes for men and women.

3 Model of College Major Choices

In this section, I introduce a generalized Roy model to capture the dynamics of major choices, educational attainment and associated labor market outcomes for students initially enrolled in four-year college. In the model, a vector of latent abilities affects educational decisions and associated labor market outcomes. The model thus follows a generalized Roy (1951) framework, fitting in with previous work by Heckman et al. (2006, 2018); Humphries et al. (2019); Rodríguez et al. (2022), allowing for individuals' choices and outcomes to depend both on their observed and unobserved characteristics.

Educational decisions are modeled sequentially, as follows: students first select an initial college major among the five options presented above. Initial major choices are unordered, as there is no natural ordering of such options. Students subsequently decide whether to continue in college or to dropout, and college continuers lastly choose whether to remain in their initial major or to switch to a different degree. Upon completing their educational attainment, students enter the labor market and after making an employment decision, earn hourly wages. I estimate the model separately for males and females to allow for differential sorting patterns by gender and to capture gender-specific labor market outcomes.

This framework combines elements from reduced form analysis and structural models to correct for endogenous educational choices and associated labor market outcomes, yet it does not postulate preferences and/or information sets, as in Arcidiacono (2004); Zafar (2013); Stinebrickner and Stinebrickner (2014); Wiswall and Zafar (2015). As such, the model does not recover the importance of belief updating and learning in the major choice process. On the other hand, I can recover the full distribution of counterfactuals — which allow me to estimate various policy-relevant treatment effects — while accounting for the importance of multidimensional skills in shaping major choices and labor market outcomes.

3.1 Model Structure

Initial Major Choice. After graduating from high school and enrolling in four-year college, students select an initial major $m \in \mathcal{M}$, where $\mathcal{M}$ encompasses the set of majors in Section 2. Their major choice depends on their observed characteristics and their latent ability ($\boldsymbol{\theta}$). Let $V_{i,m}^{G}$ be the utility for student i of gender G (male m or female f) of starting in major m.[12] $V_{i,m}$ represents an approximation of the value of each major for individual i, as it incorporates students' perceived economic returns to each major and non-pecuniary tastes. $V_{i,m}$ is given by:

$$V_{i,m} = \beta_m X_{i,m} + \alpha_m \boldsymbol{\theta_i} + \varepsilon_{i,m} \quad for \;\; m \in \mathcal{M} \tag{1}$$

where $X_{i,m}$ includes observed characteristics measured at baseline affecting major choices, $\boldsymbol{\theta_i}$ represents the vector of latent ability and $\varepsilon_{i,m}$ is an error term which is independent of observed and unobserved characteristics ($\varepsilon_{i,m} \perp X_{i,m}, \boldsymbol{\theta_i}$) as well as across major choices ($\varepsilon_{i,m} \perp \varepsilon_{i,m'}$ for $m, m' \in \mathcal{M}$). Conditional on observed characteristics and latent ability, major choices are unordered. As such, students select the college major with the highest utility:

$$D_{i,m} = \underset{m \in \mathcal{M}}{\operatorname{argmax}} \left\{ V_{i,m} \right\}.$$

Since the existing literature on college majors has previously highlighted the importance of recovering the returns to majors relative to students next-best options (Kirkeboen et al., 2016; Altonji et al., 2016), the second-best major is given by: $N_{i,j} = \operatorname{argmax}_{m \in \mathcal{M}|m^*} \left\{ V_{i,m} \right\}$ where $N_{i,j}$ is the second-best major and $\{\mathcal{M}|m^*\}$ captures the set of major choices besides the preferred choice m^*.

Final Educational Attainment. Since a sizable share of initial four-year college enrollees fail to complete a degree, the model further incorporates the college completion margin. For a student

[12] For notational simplicity, I omit the gender superscript in the rest of the Section.

i who started in major m, her decision to continue in college ($C_{i,m} = 1$) also depends on their observed characteristics and latent skills. This decision is given by:

$$C_{i,m} = \mathbb{1}\left[\beta_{m,C} X_{i,m,C} + \alpha_{m,C} \boldsymbol{\theta_i} + \varepsilon_{i,m,C} > 0\right] \tag{2}$$

where $\mathbb{1}$ is an indicator function, $\varepsilon_{i,m,C}$ is independent of observed and unobserved characteristics ($\varepsilon_{i,m,C} \perp X_{i,m,C}, \boldsymbol{\theta_i}$). $X_{i,m,C}$ encompasses observed characteristics affecting the college continuation decision. Students who dropout of college ($C_{i,m} = 0$) subsequently enter the labor market. Students who remain enrolled then decide whether to complete their initial major ($F_{i,m} = 1$) or to switch to a different degree ($F_{i,m} = 0$). This decision is given by:

$$F_{i,m} = \mathbb{1}\left[\beta_{m,F} X_{i,m,F} + \alpha_{m,F} \boldsymbol{\theta_i} + \varepsilon_{i,m,F} > 0\right] \tag{3}$$

where $\varepsilon_{i,m,F}$ is independent of observed, unobserved characteristics as well as of the error terms in equations (1)-(2). $X_{i,m,F}$ encompasses observed characteristics affecting the major switching decision. All in all, the combination of educational encompassing the choices outlined in equations (1)-(3) — given by $[D_{i,m}, C_{i,m}, F_{i,m}]$, leads to a final level of attainment $s \in \mathcal{S}$ captured by the dummy variable $D_{i,s}$.

Labor Market Outcomes. In this framework, hourly wages at age 25 represent the main labor market outcome of interest, yet wages are only observed for individuals who are employed at age 25 ($E_{i,s} = 1$). I similarly model the employment decision through the following linear specification:

$$E_{i,s} = \mathbb{1}\left[\beta_{s,E} X_{i,s,E} + \alpha_{s,E} \boldsymbol{\theta_i} + v_{i,s,E} > 0\right] \tag{4}$$

where $X_{i,s,E}$ includes the same observed characteristics previously included in the choice equations, as these variables may directly affect labor market outcomes (Heckman et al., 2018). The error term is independent of observed and unobserved characteristics. Potential hourly wages ($Y_{i,s}$) vary across students' final educational attainment and are given by the following separable specification:

$$Y_{i,s} = \beta_{s,Y} X_{i,s,Y} + \alpha_{s,Y} \boldsymbol{\theta_i} + v_{i,s,Y} \tag{5}$$

where $v_{i,s,Y}$ captures an idiosyncratic shock to hourly wages, which is independent of observed and unobserved characteristics ($v_{i,s,Y} \perp X_{i,s,Y}, \boldsymbol{\theta_i}$). Importantly, the model allows me to recover potential wages ($Y_{i,s}$) for all individuals in the analysis, regardless of whether they worked in 2011. As is standard in discrete choice models with multiple decisions (Heckman et al., 2016), I further assume that the error terms are independent across schooling decisions in equations (1)-(3), the employment decision and potential wage outcomes.[13]

While equation (5) defines wages across final levels of attainment, this parameter does not allow me to estimate the returns to initial major choices. As a result, I follow the Quandt (1958)

[13]Specifically, $v_{i,s,Y} \perp v_{i,s',Y} \perp v_{i,s,E}$ $(\forall\, s, s' \in \mathcal{S}) \perp \{\varepsilon_{i,m} \perp \varepsilon_{i,m,C} \perp \varepsilon_{i,m,F}$ $[\forall\, m, m' \in \mathcal{M}]\}$.

switching regression framework to define potential wages across initial majors $m \in \mathcal{M}$ in:

$$Y_{i,m} = Y_{i,m,F}(C_{i,m} \times F_{i,m}) + Y_{i,m,S}[C_{i,m} \times (1 - F_{i,m})] + Y_{i,m,C_0}(1 - C_{i,m}) \quad for \;\; m \in \mathcal{M} \tag{6}$$

where $Y_{i,m,F}$ represents wages for students in major m who continued in college and completed their initial major.[14] The structure of the model implies that $\boldsymbol{\theta_i}$ drives the cross-correlations of major choices and labor market outcomes. Identifying its distribution is thus of paramount importance.

3.2 Measurement System

The latent ability vector is unobserved to the econometrician, as there are no direct measures of ability available. Moreover, observed test scores measure latent abilities with error. As such, I follow an extensive literature and allow for $\boldsymbol{\theta}$ to be proxied by multiple skill measures available in the ELS. Formally, I posit a model in which observed skill measures are a linear outcome of students' latent abilities ($\boldsymbol{\theta}$) and of their background characteristics.[15]

As outlined in Section 2, I observe two measures of students' math performance, their high school grades, three non-cognitive skill measures and two math self-efficacy variables. I thus consider three components of the latent ability vector, which encompass mathematical self-efficacy (θ_{SE}), non-cognitive skills (θ_{NC}) and mathematical ability (θ_M).[16] Following the evidence presented in Figure A.1, which showed that self-efficacy measures load on a single factor whereas 'traditional' non-cognitive skill measures load on an additional factor, I allow for the self-efficacy measures to be dedicated measurements of θ_{SE} and for the non-cognitive skill measures to load on both θ_{SE} and θ_{NC}. Since the math test scores and GPA measures available in the ELS represent achievement, rather than intelligence tests, I follow Kautz et al. (2014) and allow for achievement measures to depend on math ability as well as on the factors encompassing non-cognitive abilities (θ_{NC} and θ_{SE}).[17] This measurement system allows me to assess the extent to which students' math self-efficacy affects their observed non-cognitive skill measures and their math test scores.

As such, I specify the following linear model for the two self-efficacy measures:

$$SE_{i,j} = \varphi_{SE,j} + \beta_{SE,j} X_{i,T} + \alpha_{SE,j}\theta_{i,SE} + e_{i,SE,j} \tag{7}$$

where $\boldsymbol{SE}_i$ is the vector of observed math self-efficacy measures, $\varphi_{SE,j}$ captures the intercept of self-efficacy measure j, $X_{i,T}$ is a vector of exogenous control variables encompassing students' socioeconomic status and their individual characteristics (Heckman et al., 2006, 2018). $\boldsymbol{e}_{i,SE}$ captures a mean-zero error term, independent across observed measures j, $X_{i,T}$ and θ_{SE}.

[14] $Y_{i,m,S}$ capture wages for college continuers who switch majors and Y_{i,m,C_0} indicate wages for dropouts.

[15] Since family, cultural and social factors shape the evolution of ability through childhood (Heckman, 2006), I consider the components of θ to be fixed by the time of college enrollment, but not fixed from birth.

[16] ELS data includes a measure for English self-efficacy and a reading test score. Since there is only one measure available per skill domain, a latent factor in these skill dimensions cannot be recovered (Carneiro et al., 2003).

[17] The posited structure thus implies a triangular measurement system, in which the first set of measures depends on one factor, the second set depends on the first factor along with an additional one, and so on.

I similarly posit a linear model for non-cognitive measures:

$$NC_{i,k} = \varphi_{NC,k} + \beta_{NC,k} X_{i,T} + \alpha_{SE,k}\theta_{i,SE} + \gamma_{NC,k}\theta_{i,NC} + e_{i,NC,k} \tag{8}$$

Lastly, the linear model for math test scores and GPA is given by:

$$M_{i,l} = \varphi_{M,l} + \beta_{M,l} X_{i,T} + \alpha_{SE,l}\theta_{i,SE} + \gamma_{NC,l}\theta_{i,NC} + \eta_{M,l}\theta_{i,M} + e_{i,M,l}. \tag{9}$$

where $\boldsymbol{M}_i$ encompasses both math test scores and GPA; $\boldsymbol{e}_{i,M}$ is an error term which is mutually independent from all other error terms in the model, observables $(X_{i,T})$ and latent abilities $(\boldsymbol{\theta})$.[18]

In Appendix C, I show how the measurement system secures the identification of the distribution of $\boldsymbol{\theta}$, the factor loadings $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\eta})$ and the variance of the error terms, following identification arguments introduced in Carneiro et al. (2003),Hansen et al. (2004) and Williams (2020). While triangular measurement systems like the one presented in equations (7)-(9) have previously assumed orthogonality in the latent factors (Hansen et al., 2004), orthogonal factors would imply a strong assumption in this context. By additionally restricting one of the loadings for the first factor (in this case, θ_{SE}) for one measure in each block of test scores, I can allow for the latent factors to be correlated.[19]

Gender Differences in Latent Abilities. Equations (7)-(9) are estimated separately for males and females, allowing for gender-specific parameters of the latent ability vector. Thus, for the factors to be comparable across genders, a necessary assumption is that the latent factors recovered through the measurement system presented above capture the same underlying traits for males and females. The psychometrics literature refers to this assumption as configural invariance (Kline, 2015; Putnick and Bornstein, 2016), which requires each factor to be associated with the same observed measures across groups. Configural invariance is established when the observed measures exhibit the same pattern of salient and non-salient factor loadings for men and women (Horn and McArdle, 1992).[20]

While configural invariance implies the latent factors capture the same trait for men and women, additional assumptions are needed to identify gender differences in latent abilities. I illustrate the approach using the measurement system for the self-efficacy measures. Omitting the dependence on observables for notational simplicity, taking the expectation of equation (7) by gender yields:

$$E(SE^f_{i,j}) = \varphi^f_{SE,j} + \alpha^f_{SE,j}\mu^f_{SE} \tag{10}$$

$$E(SE^m_{i,j}) = \varphi^m_{SE,j} + \alpha^m_{SE,j}\mu^m_{SE} \tag{11}$$

[18]For M_l, NC_k and SE_j, $e_{i,M,l} \perp e_{i,NC,k} \perp e_{i,SE,j} \perp v_{i,s,E} \perp v_{i,s,Y} \perp \{\varepsilon_{i,m} \perp \varepsilon_{i,m,C} \perp \varepsilon_{i,m,F} \forall\, m \in \mathcal{M}\}$.

[19]I follow the evidence presented in Figure A.1 and assume that instrumental motivation is a dedicated measure of θ_{NC} and that high school GPA does not directly depend on students' math self-efficacy. However, since these assumptions allow for the latent factors to be correlated, θ_{SE} can still affect these two observed measures indirectly.

[20]This assumption does not require the factor loadings to be the same for men and women, but rather that, for instance, if the latent self-efficacy factor has a positive and significant loading on the first self-efficacy measure for men, the corresponding factor for women loading need also be positive and significant.

where $E(\theta_{SE}^f) = \mu_{SE}^f$ and $E(\theta_{SE}^m) = \mu_{SE}^m$. Δ_{SE} denotes the average difference in self-efficacy ($\Delta_{SE} = \mu_{SE}^m - \mu_{SE}^f$). Identifying both the gender-specific factor means and the intercepts from the test score means requires additional assumptions. First, since latent factors have no location of their own, we can normalize the mean of the latent factor for one group ($\mu_{SE}^f = 0$ without loss of generality). The difference in the test score means becomes:

$$E(SE_{i,j}^m) - E(SE_{i,j}^f) = (\varphi_{SE,j}^m - \varphi_{SE,j}^f) + \alpha_{SE,j}^m \Delta_{SE}. \tag{12}$$

Identifying Δ_{SE} from equation (12) requires an additional assumption. In particular, the test score intercepts must be equivalent for men and women ($\varphi_{SE,j}^m = \varphi_{SE,j}^f$). This assumption is referred to as scalar invariance, and it requires that there are no consistent cross-group differences in how each group performs/answers the observed measures.[21] Under this assumption, group differences in the observed measure means will thus reflect gender gaps in the factor means. Δ_{SE} is thus directly identified as: $\Delta_{SE} = [E(SE_{i,j}^m) - E(SE_{i,j}^f)]/\alpha_{SE,j}$.

The same identification argument follows for the identification of gender gaps in non-cognitive skills. Normalizing $\mu_{NC}^f = 0$ implies the average gender gap in observed non-cognitive skill measure k (equation (8)) is given by:

$$E(NC_{i,k}^m) - E(NC_{i,k}^f) = (\varphi_{NC,k}^m - \varphi_{NC,k}^f) + \alpha_{SE,j}^m \Delta_{SE} + \gamma_{NC,k}^m \Delta_{NC} \tag{13}$$

where Δ_{NC} captures gender differences in the mean of θ_{NC}. Scalar invariance implies that Δ_{NC} is identified as the remaining parameters in equation (13) are already identified. The same identification argument follows for math ability, as discussed in detail in Appendix D. Importantly, when the configural and scalar invariance assumptions hold, group differences in the latent factors are identified. The estimated factors can be thus compared across males and females (Gregorich, 2006).

3.3 Identification and Estimation

Model Identification. Carneiro et al. (2003), Heckman and Navarro (2007) and Heckman et al. (2016) present the formal argument for identification of a multi-stage sequential choice model, akin to the one presented in this paper. The distribution of latent ability $\boldsymbol{\theta}$ is identified through the measurement system in equations (7)-(9), which requires for $\boldsymbol{\theta}$ to be orthogonal to $\boldsymbol{X}$ and ε. While data availability implies that all observed measures are available for 60% of the sample, Williams (2020) shows that the distribution of the latent factors is identified as long as the variance-covariance matrix of observed measures can be consistently estimated.[22] Furthermore, Hansen et al. (2004) Heckman et al. (2016) show that in the absence of exclusion restrictions, the joint distribution of choices and potential outcomes can be non-parametrically identified as long as the support on the

[21] This assumption would not hold in the case in which one group has a tendency to systematically report higher (or lower) answers to questions (Gregorich, 2006). Table A.1 had shown that gender differences in self-efficacy only emerge in mathematics, making it unlikely that such tendencies could emerge in this context.

[22] Despite small differences in non-respondents' observed characteristics (Table A.2), Piatek and Pinger (2016) show the parameters in the choice equations should be equivalent for individuals with and without missing measures.

covariates in the choice equations (for instance, $\beta_m X_{i,m}$ in equation (1)) matches the support of the corresponding error terms ($\psi_m = \alpha_m \boldsymbol{\theta_i} + \varepsilon_{i,m}$). In this context, a conditional independence assumption — which implies that initial major choices, subsequent educational choices and labor market outcomes are independent conditional on all observed characteristics and latent ability (a 'matching-on-unobservables assumption') — secures model identification.

Model Implementation. The distribution of $\boldsymbol{\theta}$ is identified non-parametrically (Freyberger, 2018), yet for computational convenience, I estimate the density of each unobserved ability component f for each gender by using a mixture of three normal distributions with means $(\mu_{1,f}, \mu_{2,f}, \mu_{3,f})$, probabilities $(p_{1,f}, p_{2,f}, p_{3,f})$, with $p_{1,f} + p_{2,f} + p_{3,f} = 1$, and variances $((\sigma_{1,f})^2, (\sigma_{2,f})^2, (\sigma_{3,f})^2)$ as follows:

$$\theta_f \sim p_{1,f}\ N(\mu_{1,f}, (\sigma_{1,f})^2) + p_{2,f}\ N(\mu_{2,f}, (\sigma_{2,f})^2) + p_{3,f}\ N(\mu_{3,f}, (\sigma_{3,f})^2)$$

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 observed test scores and non-cognitive skill measures $t \in \mathcal{T}$ in $\mathbf{T}_i$. Let Ψ be the vector of model parameters.[23] While the model is identified non-parametrically, I estimate the model using normal distributions for the idiosyncratic shocks in the measurement system, educational choice equations, employment decision and in the wage equation.[24] Given the independence assumptions invoked above, the likelihood for a set of I individuals is given by:

$$\begin{aligned}\mathcal{L}(\Psi \mid \cdot) \quad = \quad & \prod_{i \in \mathcal{I}} \Bigg[\int_{\boldsymbol{\theta}} \prod_{t \in \mathcal{T}} f(T_{it} \mid X_{iT}, \boldsymbol{\theta}) \prod_{s \in \mathcal{S}} \Big\{ P(D_{is}{=}1 \mid \mathbf{X}_i, \boldsymbol{\theta}) \left[f(Y_{is} \mid X_{isY}, \boldsymbol{\theta}) P(E_{is}{=}1 \mid X_{isE}, \boldsymbol{\theta}) \right]^{E_{is}} \\ & [1 - P(E_{is} = 1 \mid X_{isE}, \boldsymbol{\theta})]^{1-E_{is}} \Big\}^{D_{is}} dF_{\boldsymbol{\theta}}(.) \Bigg]\end{aligned}$$

where $f(T_t|\cdot)$ is the conditional density function of test score t, $f(Y_s|\cdot)$ is the conditional density function of hourly wages for schooling level s and $F(\boldsymbol{\theta})$ represents the cumulative distribution function of the latent factors. I estimate the model using a Gibbs sampler as the Markov Chain Monte Carlo (MCMC) algorithm, as in Hansen et al. (2004); Heckman et al. (2006).[25] I generate 500 draws from the estimated posterior distribution of the model parameters and simulate 200 samples

[23] The measurement system, educational choice equations and labor market outcomes include students' race, family composition, parents' education and household income as control variables ($\mathbf{X}_i$). Factor loadings in wage equations with fewer than fifteen individuals – female STEM dropouts and male life science dropouts – are restricted to zero.

[24] Specifically, $e_{i,SE,j} \sim N(0, \sigma^2_{SE_j}) \forall j \in \mathcal{J}$; $e_{i,NC,k} \sim N(0, \sigma^2_{NC_k}) \forall k \in \mathcal{K}$; $e_{i,M,l} \sim N(0, \sigma^2_{M_l}) \forall l \in \mathcal{L}$; $[(\varepsilon_{i,m} \sim N(0,1); \varepsilon_{i,m,C} \sim N(0,1); \varepsilon_{i,m,F} \sim N(0,1)) \forall m \in \mathcal{M}]$; $v_{i,s,E} \sim N(0,1) \forall s \in \mathcal{S}$; $v_{i,s,Y} \sim N(0, \sigma^2_{s,Y}) \forall s \in \mathcal{S}$. The initial major choice decision is estimated with a multinomial probit. Equations (2)-(4) are estimated using a probit model.

[25] Using a vector of initial parameters from the transition kernel, the Markov Chain is generated according to the Gibbs sampler, whose limiting distribution is the posterior. Once convergence is achieved and after a burnin period of 100 draws, I keep every thirtieth draw to generate a sample of 500 draws from the posterior distribution of estimated model parameters to compute the mean and the standard errors of the parameters of interest. Appendix E describes the estimation algorithm in detail.

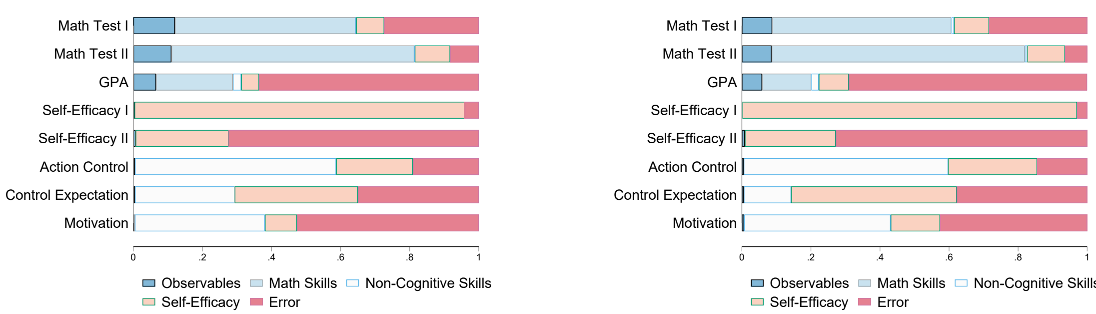
where each simulated sample draws from the posterior of the estimated model parameters. Inference follows standard Bayesian arguments, as the Bernstein-von Mises theorem allows me to obtain the associated standard errors from the standard deviations computed from these draws.

Goodness of Fit. Figures A.2-A.5 present evidence on the goodness of fit of the model. The model accurately predicts major choices by gender (Figure A.2), and also matches the share of students who complete their initial major (Figure A.3). Moreover, the model closely matches employment rates and hourly wages for initial major completers across their parents' income (Figures A.4-A.5).

4 Latent Skills and College Major Choices

4.1 Measurement System and Latent Skills

Figure 2: Variance Decomposition of Measurement System by Gender



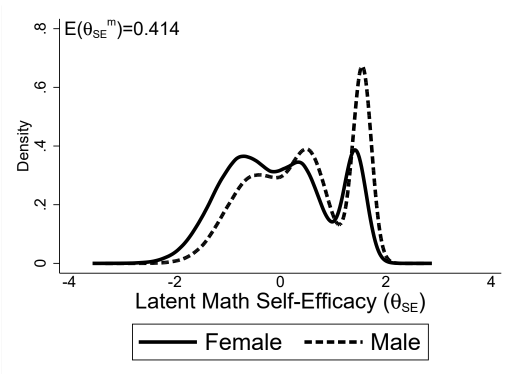
(a) Variance Decomposition: Women

(b) Variance Decomposition: Men

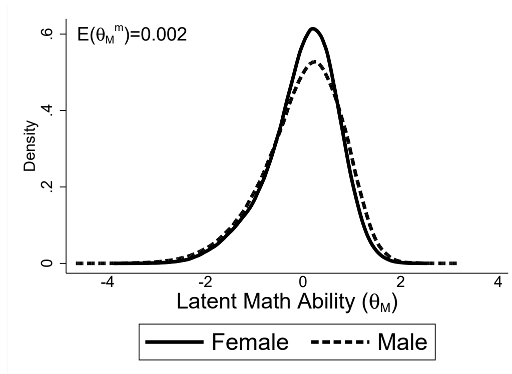
Note: Figure 2 presents the contribution of observed characteristics, latent abilities ($\boldsymbol{\theta}$) and the error term to the variance of the observed skill measures considered in the model. The row Observables indicates the share of the variance of the measurement variables explained by background characteristics. Each Ability bar indicates the share of the variance explained by each component of the latent ability vector. The Error row represents the share of each test score variance explained by the unobserved idiosyncratic error of the measurement system. The first panel presents results for women and the second panel presents evidence for men.

Measurement System. To understand the relative contribution of students' background characteristics and their latent ability vector to each test score, I present a variance decomposition of the measurement system in Figure 2. For both math test scores and high school grades, the share explained by observable characteristics is close to 10 percent for both men and women. Observed characteristics explain a much smaller share of the variance in students' reported math self-efficacy and their non-cognitive skill measures. On the other hand, this exercise confirms the critical role of latent ability for explaining the variance in the observed measures. Across both math assessments, students' latent math ability explains between 52 and 73% of the variance in performance. Meanwhile, a sizable share of the variance in observed self-efficacy measures is explained by their latent self-efficacy. Lastly, 38-59% of the variance in students' action control and motivation is explained by θ_{NC}, yet an additional share is explained by their math self-efficacy. All in all, this evidence supports the argument that test scores cannot be equated with ability, as they are direct functions of background characteristics and capture distinct components of the ability vector. The same factor loadings are salient in each test score equation for females and males (Tables F.1-F.2), and the share of the test score variance explained by each ability component is similar by gender, lending support to the configural invariance assumption imposed in Section 3.2.

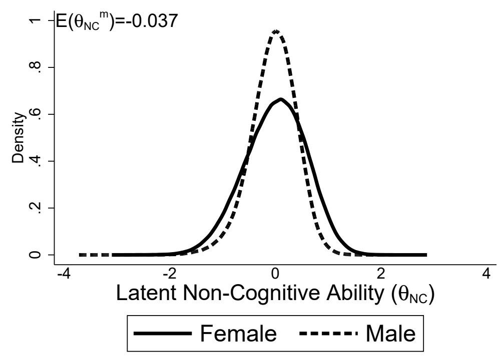
Figure 3: Marginal Distributions of Latent Ability Factors by Gender



(a) Density of θ_{SE}



(b) Density of θ_M



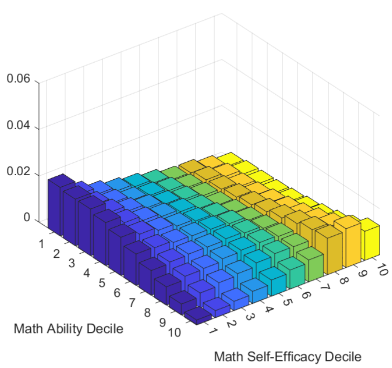
(c) Density of θ_{NC}

Note: Figure 3 shows the marginal densities of the estimated latent math ability, math self-efficacy and non-cognitive skill factors for men and women. The mean of the three latent factors is normalized to equal zero for women (see Section 3.2), and the estimated mean for each latent factor for males is reported in each panel. The parameters of the latent factors are presented in the first panel of Table A.6. For the three latent factors, Kolmogorov-Smirnov tests of equality between the male and female distributions are rejected with p-values smaller than 0.001.

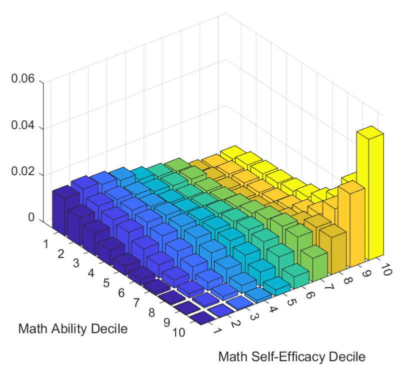
Latent Skills. Figure 3 presents kernel density estimates of the marginal distributions of the latent ability factors by gender. The first panel presents the estimated distribution of latent math self-efficacy for men and women. Following the identification arguments introduced in equations (10)-(12), men outpace women in math self-efficacy by 0.41 standard deviations, on average. Moreover, a Kolmogorov-Smirnov test of the equality of the male and female distributions is rejected with a p-value smaller than 0.001. The next two panels present the corresponding distributions of students' latent non-cognitive ability and their math ability, respectively. In this case, the average the latent factors exhibit similar means for females and males, such that the average gender gap in these two skill dimensions is not statistically significant (Table A.6). Altogether, the distributions of all three factors are different for men and women, yet the only statistically significant gender gap in average skills emerges for students' math self-efficacy.

I find a large and positive correlation across the three latent factors for men and women. Most important to the analysis of sorting into STEM, however, is the correlation between the students' latent math ability and their self-efficacy. Sizable differences emerge in this dimension, as the estimated correlation between θ_M and θ_{SE} for men equals 0.488, far surpassing the 0.304 correlation for women (Table A.6). Importantly, under configural and scalar invariance, the latent ability factors represent the same underlying traits for men and women. I can thus construct a distribution for each ability factor that encompasses both men and women, allowing me to, for instance, assess the share of women that belong to the top math self-efficacy decile of the joint distribution.[26] Since women's math self-efficacy is substantially lower than that of men, a far lower share of high math-skilled women have high math self-efficacy vis-à-vis their male counterparts. Figure 4 presents the joint distribution of math ability and self-efficacy. While 43.3% of men in the top math ability decile are also in the top math self-efficacy decile, just 13.9% of high-math-skilled women reach the top θ_{SE} decile.[27] Altogether, there is a large under-representation of high-skilled women who exhibit strong confidence in their math abilities.[28]

Figure 4: Joint Distribution of Latent Math Ability and Self-Efficacy by Gender



(a) Joint Distribution of θ_M and θ_{SE}: Women



(b) Joint Distribution of θ_M and θ_{SE}: Men

Note: Figure 4 presents the joint distribution of math ability and self-efficacy. The math ability and self-efficacy deciles are given by the joint distribution encompassing men and women, presented in the third column of Panel A of Table A.6. The figure documents the share of women (Panel A) and men (Panel B) pertaining to each joint decile of the two latent factor distributions.

[26] The parameters of the combined distribution are presented in the third column of the first panel in Table A.6.

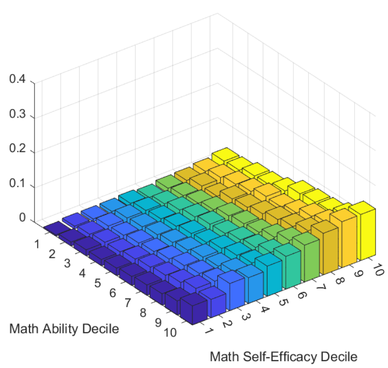
[27] These results hold even when examining the gender-specific distribution of math ability and self-efficacy (Figure A.6), as a lower share of high math ability women are in the top decile of their own-gender's θ_{SE} distribution.

[28] External influences may drive the prevalence of high-skilled women who lack confidence in their math ability. Carlana (2019) finds that teachers' gender stereotypes lower girls' subsequent performance and self-confidence in math, Lavy and Sand (2018) show similar evidence on the impact of teachers on female students' test scores in Israel.

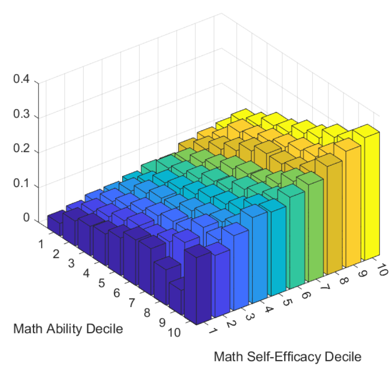
4.2 Initial Major Choices

In light of the relative lack of high math ability women with high math self-efficacy, I examine the relationship between students' latent skills and their STEM enrollment decisions in Figure 5. The first panel shows that women who are in the top joint decile of the math ability and self-efficacy distribution are far more likely to start in STEM (13.3 percent) than those in the middle joint decile (3.9 percent). Self-efficacy plays a critical role in this decision: among women in the top math ability decile, moving from the bottom self-efficacy decile to the top one increases STEM participation rates by 7.8 percentage points. The second panel presents evidence for men, which shows that 32.9% of men in the top decile of the math ability distribution enroll in STEM, far outpacing women's STEM enrollment rates. For men in the top θ_M decile, moving from the bottom decile of θ_{SE} to the top one would almost double their STEM participation, from 19.4% to 35.4%.[29] For students in the top joint decile of the math ability and self-efficacy distribution, a similar share of men and women enroll in life science and business majors (11-13%). The largest gender differences emerge in health fields (10.6% for women, 2.9% for men) and in 'Other' majors, where female participation equals 51.4% compared to 38.2% for men.

Figure 5: STEM Enrollment Rates by θ_M and θ_{SE} by Gender



(a) STEM Enrollment Rates: Women



(b) STEM Enrollment Rates: Men

Note: Figure 5 shows the share of women and men who initially enroll a STEM degree. The share of STEM enrollees is presented across each decile of the joint math ability and self-efficacy distribution, where the deciles are defined by the joint distribution of ability encompassing both men and women, presented in the third column of Panel A of Table A.6. Figure A.8 presents corresponding evidence of STEM enrollment by the gender-specific joint deciles of math ability and self-efficacy.

These results thus show that despite similar sorting-into-STEM patterns for men and women, a far higher share of high-math-ability men start in STEM majors vis-à-vis their female counterparts. As such, these findings fit in with an extensive literature showing that STEM gaps are largely driven by gender differences in preferences and tastes (Zafar, 2013; Wiswall and Zafar, 2015, 2018; Reuben et al., 2017; Patnaik et al., 2020). At the same time, the evidence presented in Figure 5 shows that

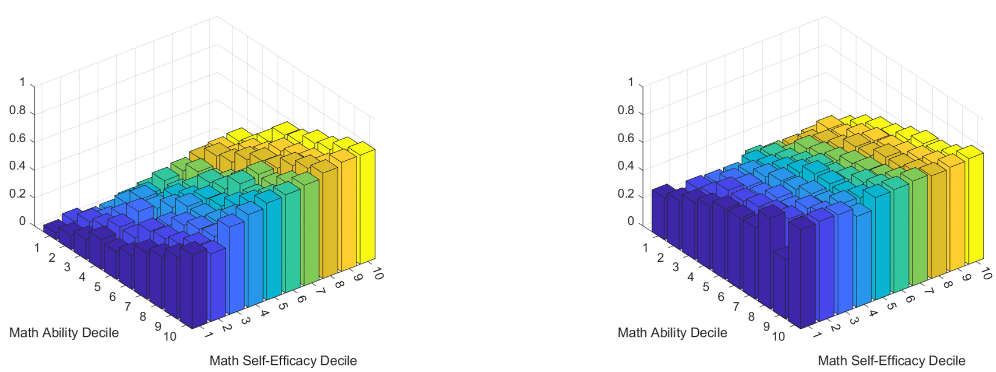
[29]Figure A.7 shows how students sort into majors based on their latent abilities. Panels (a)-(d) show that STEM enrollees have the highest average math ability and self-efficacy, followed by their peers in life sciences and business. Panels (e)-(f) show there are far smaller differences in students' non-cognitive skills across majors. Table F.3 (Panel A) presents the estimated factor loadings in the major choice equation, testing for gender differences in the loadings.

increasing women's math self-efficacy could increase their STEM participation.

4.3 Final Major Choices

While 62% of male STEM enrollees end up successfully completing a major in this field, this is the case for just 55% of their female counterparts. In Figure 6, I further examine how math ability and self-efficacy jointly affect the likelihood of STEM completion for students who started in these majors. The first panel shows that both math ability and self-efficacy significantly strongly shape the likelihood of STEM degree completion for women. For instance, moving from the middle to the top decile of the marginal math ability distribution increases the likelihood of STEM completion from 37.7% to 74.1%. Math self-efficacy is similarly important, as the corresponding increase from the middle to the top decile of the θ_{SE} distribution would boost completion rates from 47.3% to 68.1%. The joint skill distribution of math ability and self-efficacy presents a similar story, as 78.5% of women in the top joint decile graduate with a STEM degree, yet this share drops to below 54% percent for those in the top math ability decile and in the bottom of the θ_{SE} distribution.

Figure 6: STEM Completion Rates Among Initial Enrollees: by Gender



(a) STEM Completion Rates: Women

(b) STEM Completion Rates: Men

Note: Figure 6 shows the share of women and men who complete a STEM degree after initially enrolling in a math-intensive major. The share of conditional completers is presented across each decile of the joint math ability and self-efficacy distribution, where the deciles are defined by the joint distribution of ability encompassing both men and women, presented in the third column of Panel A of Table A.6. The first panel presents results for women. The second panel presents evidence for men. Figure A.9 presents corresponding evidence of STEM completion by the gender-specific joint deciles of math ability and self-efficacy.

The second panel of Figure 6 shows that these patterns are strikingly different for men, for whom self-efficacy plays a far smaller role in driving STEM degree completion. While 56% of STEM enrollees in the middle decile of the θ_{SE} distribution complete a degree after enrollment, this share rises only slightly 67.7% top self-efficacy decile. Among men in the top decile of the math ability distribution, self-efficacy similarly plays a limited role in shaping the likelihood of successfully completing a STEM degree. This result shows the importance of considering how different margins of ability differentially affect men and women's progress through majors in college.[30] Specifically,

[30]The findings presented in Figures 4-6 hold in an alternative version of the model presented in Section 3 which

math self-efficacy plays a critical role for women's exit from STEM, yet this margin has not received much attention in the existing literature.[31]

While the model presented in Section 3 does not directly account for the mechanisms through which low math self-efficacy drives STEM dropout, women with low self-efficacy may leave STEM upon receiving a low grade in a STEM-based class. In this context, Rask and Tiefenthaler (2008) and Goldin (2015) have shown evidence pointing in this direction across two liberal arts colleges. Women who earn low grades in introductory economics courses are more likely to dropout from economics majors than their male counterparts. Kugler et al. (2021) similarly find that women enrolled in math-intensive STEM fields are more likely to leave such majors in response to low grades than their male peers. Ahn et al. (2022) show that harsh grading policies in STEM majors may exacerbate the gender gap in these majors, as women place a higher value on grades than men. Altogether, women's lack of math self-efficacy may influence their large STEM dropout rates through the grades received in introductory courses.

5 Returns to College Majors

5.1 Conceptual Framework

While STEM-promoting policies may create important non-pecuniary benefits, understanding the wage returns associated with these majors is a first-order concern for quantifying the benefits arising from such interventions. In this context, an extensive literature has estimated the returns to completing different majors (Altonji, 1993; Rumberger and Thomas, 1993; Chevalier, 2011; Webber, 2014; Jiang, 2021), yet the evidence regarding the gender-specific returns to enrolling in different majors has been so far limited.[32] I take advantage of the estimated model parameters and the potential wages across initial majors defined in equation (6) to recover the returns to STEM majors relative to various alternative options. The estimated returns to majors only capture early-career wage outcomes, thus not recovering the lifecycle returns to majors (Altonji et al., 2016), nor potential non-pecuniary benefits from such choices (Oreopoulos and Salvanes, 2011).

Let E[.] denote the expected value taken with respect to the distribution of (X, $\boldsymbol{\theta}$, ε). The average treatment effect (ATE) of enrolling in a STEM major (S) relative to any other major ($m \in \mathcal{M}$) is given by:

$$ATE^G_{S,m} \equiv \int\int E[Y^G_{i,S} - Y^G_{i,m} | \boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}] dF_{X,\boldsymbol{\theta}}(x, \underline{\boldsymbol{\theta}}) \;\; for \;\; m, f \in \{G\} \tag{14}$$

where $Y^G_{i,S} - Y^G_{i,m}$ captures the wage returns to starting in a STEM major relative to enrolling in

incorporates non-college-enrollees and directly models students' choice of initial enrollment levels (Figures B.1-B.3).

[31] For STEM enrollees, non-cognitive ability does not lead to an increased likelihood of STEM completion (Figure A.10). For non-STEM enrollees, math ability significantly reduces the likelihood of college dropout, and the other two skill dimensions play a smaller role in shaping completion outcomes (Figure A.11). The last two panels of Table F.3 present the estimated factor loadings in the college continuation and major completion equations.

[32] Tables F.4-F.5 present the estimated factor loadings in the employment and wage equations. The estimated factor loadings are largely similar by gender, although a few statistically significant differences emerge.

major m for student i of gender G. The average returns to STEM enrollment are computed by integrating out the latent skill distribution, yet may be heterogeneous across $\boldsymbol{\theta}$, as latent skills may influence the likelihood of college graduation and may improve labor market productivity. I examine the heterogeneous returns to STEM enrollment across the latent ability distribution in:

$$ATE^G_{S,m}(\underline{\theta}_M, \underline{\theta}_{SE}, \underline{\theta}_{NC}) \equiv E[Y^G_{i,S} - Y^G_{i,m} | \theta_C = \underline{\theta}_C, \theta_{SE} = \underline{\theta}_{SE}, \theta_{NC} = \underline{\theta}_{NC}]. \tag{15}$$

5.2 Estimated Returns to Majors

In the first panel of Table 3, I present the estimated returns to enrolling in a STEM major for women vis-à-vis the other four majors considered in the analysis. Using the simulated parameters from the model, the second column of Table 3 shows that the estimated ATE of STEM enrollment relative to the life sciences equals 11.5%. The estimated returns to STEM exhibit further heterogeneity depending on which major they are compared against. For instance, while the average returns to STEM enrollment relative to starting in 'Other' majors are small, the returns relative to business majors are large and significant, equaling -11.4%. The returns to enrolling in STEM compared to a health-related field are even smaller, reaching -24%, yet these returns are calculated in the early-career when health-based graduates are among the highest-earners, a pattern which does not remain through the lifecycle (see Figure A.12). The estimated ATE of STEM enrollment relative to the other majors is lower when compared to average raw wage gaps across fields. This pattern is driven by the fact that women enrolled in STEM outpace their peers in other majors in both their latent math ability and self-efficacy.

The first row of the second panel of Table 3 presents the corresponding returns to STEM enrollment for men. Enrolling in STEM delivers positive average returns relative to majors in the life sciences, business and 'Other' fields, while yielding negative returns compared to a health-based field. In fact, the estimated ATE of STEM enrollment for men exceeds 22% when compared to the life sciences or other majors, fitting in with previous work (Eide, 1994; Chevalier, 2011; Jiang, 2021) which had previously found higher returns to STEM majors for men than for women.

The estimated returns may further differ for students who actually enrolled in STEM ($D^G_{i,S} = 1$) – for whom the relevant returns are captured by the treatment on the treated (TT) parameter – and those who instead enrolled in major m ($D^G_{i,m} = 1$), whose relevant returns to STEM enrollment are given by the treatment on the untreated parameter (TUT). The first panel of Table 3 shows the TT and TUT parameters for women are largely similar in magnitude to the estimated ATEs presented above. For men, the estimated TT to STEM enrollment is larger than the corresponding ATE across all alternative majors, except with respect to health.[33]

While the returns to majors discussed so far are computed by integrating out the latent skill distribution, these may be heterogeneous across students' abilities. Table 3 further presents the returns to STEM enrollment for students above and below the math ability median, showing that

[33]Table A.7 decomposes wage differences across majors into the estimated ATE, selection bias and 'sorting gains,' which capture the difference in the TT and ATE (Heckman et al., 2018). Women are positively selected into STEM, yet selection bias plays a smaller role for men. For both genders, the sorting gains parameter is small in magnitude.

the returns to STEM for high-skilled men and women are generally larger than the corresponding average treatment effects.[34] These findings thus fit in with a higher productivity of math ability for male and female students in math-intensive degrees (Arcidiacono, 2004; Humphries et al., 2019). At the same time, previous work has found a low elasticity of college major choices to expected earnings (Berger, 1988; Beffy et al., 2012; Wiswall and Zafar, 2015; Blom et al., 2021), and since preferences play a key role in shaping major choices (Zafar, 2013; Wiswall and Zafar, 2015, 2018), high-skilled women may not act upon the positive returns from STEM majors.

Table 3: Estimated Returns to STEM v. Alternative Majors by Gender

		Panel A. Women		
Estimate	Life Sciences	Business	Health	Other
ATE	0.115	-0.114	-0.239	-0.004
	[0.112,0.118]	[-0.117,-0.112]	[-0.241,-0.236]	[-0.007,-0.002]
ATE (High θ_M)	0.154	-0.070	-0.243	0.041
	[0.150,0.158]	[-0.073,-0.066]	[-0.247,-0.240]	[0.038,0.044]
ATE (Low θ_M)	0.067	-0.168	-0.233	-0.059
	[0.062,0.071]	[-0.172,-0.164]	[-0.237,-0.229]	[-0.063,-0.056]
TT	0.111	-0.103	-0.244	-0.002
	[0.098,0.124]	[-0.114,-0.091]	[-0.256,-0.232]	[-0.014,0.009]
TUT	0.091	-0.083	-0.223	-0.007
	[0.080,0.102]	[-0.091,-0.076]	[-0.230,-0.217]	[-0.010,-0.004]
		Panel B. Men		
Estimate	Life Sciences	Business	Health	Other
ATE	0.227	0.071	-0.111	0.229
	[0.224,0.230]	[0.068,0.073]	[-0.115,-0.107]	[0.227,0.232]
ATE (High θ_M)	0.294	0.134	-0.131	0.261
	[0.290,0.298]	[0.130,0.137]	[-0.136,-0.126]	[0.258,0.264]
ATE (Low θ_M)	0.147	-0.004	-0.087	0.191
	[0.143,0.152]	[-0.008,0.000]	[-0.094,-0.081]	[0.187,0.195]
TT	0.272	0.110	-0.128	0.241
	[0.266,0.279]	[0.104,0.115]	[-0.137,-0.119]	[0.235,0.246]
TUT	0.241	0.065	-0.096	0.224
	[0.229,0.252]	[0.059,0.071]	[-0.118,-0.074]	[0.221,0.228]

Notes: Table 3 presents the estimated returns to STEM enrollment relative to different majors for women (Panel A) and for men (Panel B). The estimated average treatment effect is defined in equation (14). ATE (High θ_M) and ATE (Low θ_M) present the estimated ATE to STEM majors for students above and below the θ_M median. The TT and TUT parameters encompass the returns to individuals who actually enrolled in STEM and to those who enrolled in the alternative major under consideration, respectively. 95% confidence intervals are presented in brackets.

Returns Relative to Next-Best Option. While the results presented so far indicate that the returns to STEM are heterogeneous depending on the major they are compared against, students may only weigh the costs and benefits of their preferred and next-best majors, rather than considering all available majors. As such, the returns presented above may not represent an actionable margin for students (Rodríguez et al., 2016). To this end, I follow Heckman et al. (2008) and Kirkeboen et al. (2016) to instead examine the returns to STEM majors for students who enrolled

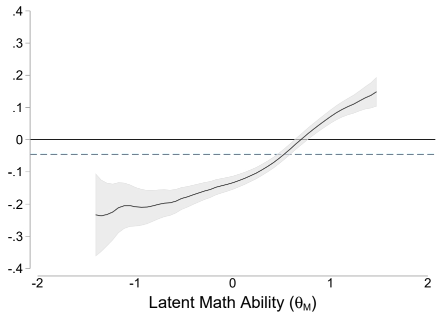
[34] Figures A.13-A.14 present heterogeneous returns to STEM majors for women and men across the θ_M distribution.

in such majors vis-à-vis their next-best option (defined in Section 3.1) as follows:

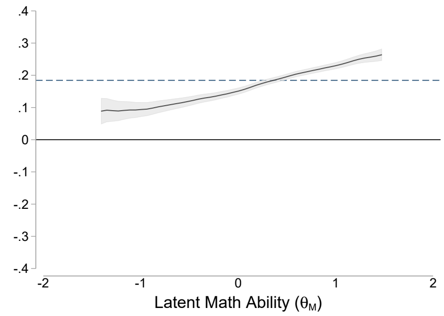
$$TT_{S,j} \equiv \int\int E[Y_{i,S} - Y_{i,j} | \boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}, D_{i,S} = 1, N_{i,j} = 1] dF_{X,\boldsymbol{\theta}}(x, \underline{\boldsymbol{\theta}}, D_{i,S} = 1, N_{i,j} = 1) \quad (16)$$

where $Y_{i,S} - Y_{i,j}$ recovers the returns to STEM enrollment compared to major j. $TT_{S,j}$ denotes the returns to STEM majors for students enrolled in STEM ($D_{i,S} = 1$) whose second-best option is major j ($N_{i,j} = 1$).[35]

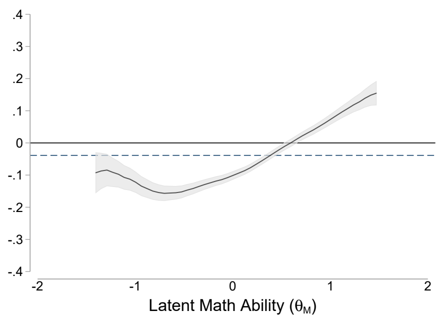
Figure 7: Heterogeneous Returns to STEM Enrollment Compared to Second-Best Option



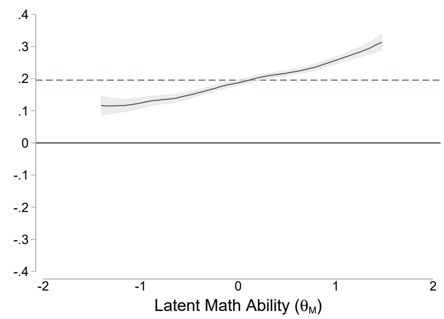
(a) TT Parameter: Women



(b) TT Parameter: Men



(c) TUT Parameter: Women



(d) TUT Parameter: Men

Note: The first panel of Figure 7 presents heterogeneous returns to STEM enrollment for women who actually enrolled in STEM relative to the next-best major (defined in equation (16)) across the θ_M distribution. The second panel presents corresponding returns for men. The third and fourth panels present the estimated TUT parameter which, for students enrolled in other majors, recovers the returns to instead enrolling in STEM for men and women, respectively. Dashed lines in each panel denote the average TT/TUT for each gender. Shaded areas denote 95% confidence intervals.

In Figure 7, I present the estimated returns to STEM relative to students' next-best options. The dashed line in the first panel shows that for women in STEM, enrolling in their next-best option would have led to an average wage gain of 4.5%. At the same time, these returns are highly heterogeneous across the θ_M distribution — $TT_{S,j}$ is positive and significant for women in the top decile of the math ability distribution, denoting that these students would benefit from remaining in STEM. For men, meanwhile, the estimated returns indicate that remaining enrolled in STEM

[35]For both female and male STEM enrollees, enrolling in a major in the 'Other' grouping would have been the preferred alternative for over half of the sample (Table A.8).

instead of their next-best major would have been beneficial for all students, with increasing returns across the math ability distribution, as well.

The corresponding treatment-on-the-untreated parameter ($TUT_{S,j}$) captures whether students enrolled in other majors m whose next best-major was in STEM would have enjoyed positive benefits from alternatively enrolling in STEM. The third panel of Figure 7 presents the estimated $TUT_{S,j}$ parameter, showing that for women enrolled in other majors, enrolling in their next-best choice (STEM) would have, on average, resulted in average wage losses of 3.9%. However, this parameter becomes positive for female students in the top quintile of the math ability distribution, fitting in with the positive returns to STEM enrollment for high math ability women shown above.

5.3 Conditional Returns to Degree Completion

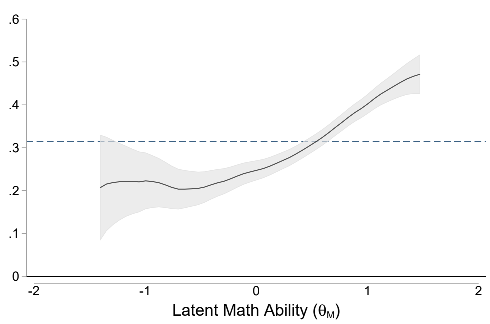
While STEM enrollment for high-math-ability women is largely positive, an open question remains as to whether remaining in these majors through completion would be necessarily deliver positive wage returns. I define the conditional returns to completing initial major m relative to switching to a different major, or relative to dropping out of college altogether as follows:

$$\{ATE_{m,F,S}|D = m\} \equiv \int\int E[Y_{i,m,F} - Y_{i,m,S}|\boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}, D = m]dF_{X,\boldsymbol{\theta}|D=m}(x, \underline{\boldsymbol{\theta}}) \tag{17}$$

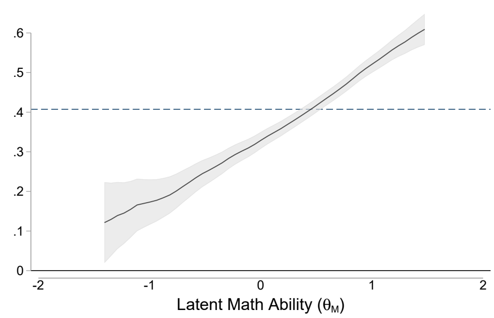
$$\{ATE_{m,F,C_0}|D = m\} \equiv \int\int E[Y_{i,m,F} - Y_{i,m,C_0}|\boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}, D = m]dF_{X,\boldsymbol{\theta}|D=m}(x, \underline{\boldsymbol{\theta}}) \tag{18}$$

where for students initially enrolled in major m, $ATE_{m,F,S}|\{D = m\}$ captures the returns to completing that major relative to switching to a different degree and $ATE_{m,F,C_0}|\{D = m\}$ denotes the returns to initial-major completion compared to dropping out of college altogether.

Figure 8: Heterogeneous Returns to STEM Completion for Female STEM Enrollees



(a) Returns v. Switching



(b) Returns v. Dropout

Note: The first panel of Figure 8 presents heterogeneous returns to STEM completion relative to switching to a different major (equation (17)) for women. The second panel presents corresponding evidence relative to dropping out of college altogether (equation (18)). The dashed lines depict the corresponding average treatment effect in each panel. Shaded areas denote 95% confidence intervals.

In the first and second panel of Figure 8, I present the estimated returns to completing a

STEM degree relative to switching to a different major and to dropping out of college, respectively. The average returns to STEM completion vis-à-vis switching are large and significant, reaching 31.5%. Similar to the returns to STEM enrollment, these returns are also increasing across the θ_M distribution, denoting that high-ability women would have the most to gain by completing these degrees. The returns relative to college dropout are larger (41%) and increasing in the math ability distribution. Men would similarly enjoy sizable returns for men from completing STEM degrees both relative to switching to a different major (36%) and to dropping out of college altogether (65%) (Figure A.15). Altogether, these findings indicate that well-targeted policies aimed at increasing STEM completion among female enrollees would substantially improve their early-career wages.

6 Policy Simulation: Math Self-Efficacy Increase

6.1 Simulated Intervention

Colleges across the country have implemented policies aimed at boosting students' STEM participation rates, ranging from mentoring initiatives, STEM-program exposure, increased lab experience and summer preparation programs (Olson and Riordan, 2012). In this context, previous work has shown that non-cognitive skills are malleable through adolescence (Kautz et al., 2014), and the psychology literature has found different strategies that can boost students' self-efficacy (Betz and Schifano, 2000; Siegle and McCoach, 2007; Cordero et al., 2010).[36] Since this literature has not been precise about the feasibility of interventions of varying magnitudes, I examine how a program which would boost women's math self-efficacy by 0.25 standard deviations — smaller than the average gender gap in self-efficacy — would affect their STEM participation rates. I also assess the effect of larger and smaller increases in self-efficacy. In light of low enrollment rates among high-math-ability women, I focus the simulated intervention on students above the math ability median. The intervention is presented for illustrative purposes, as the sustained self-efficacy increases of this magnitude may not necessarily be achieved in a large scale.

Conceptual Framework. To fix ideas, I follow the potential outcomes framework to capture the effect of the simulated intervention 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: STEM-always-takers and STEM-never-takers. The effect of any policy p' on outcome Y is given by:

$$
\begin{aligned}
\Delta^Y = E[Y_i(p') - Y_i] = \\
& E[Y_i(p') - Y_i | D_{i,s}(p') = 1, D_{i,s} = 1] \times \underbrace{P[D_{i,s}(p') = 1, D_{i,s} = 1]}_{STEM\ Enrollment\ Always-Takers} + \\
& E[Y_i(p') - Y_i | D_{i,s}(p') = 1, D_{i,s} = 0] \times \underbrace{P[D_{i,s}(p') = 1, D_{i,s} = 0]}_{STEM\ Enrollment\ Compliers} +
\end{aligned}
$$

[36] Alan et al. (2019) show that a classroom-based early childhood intervention can foster students' grit.

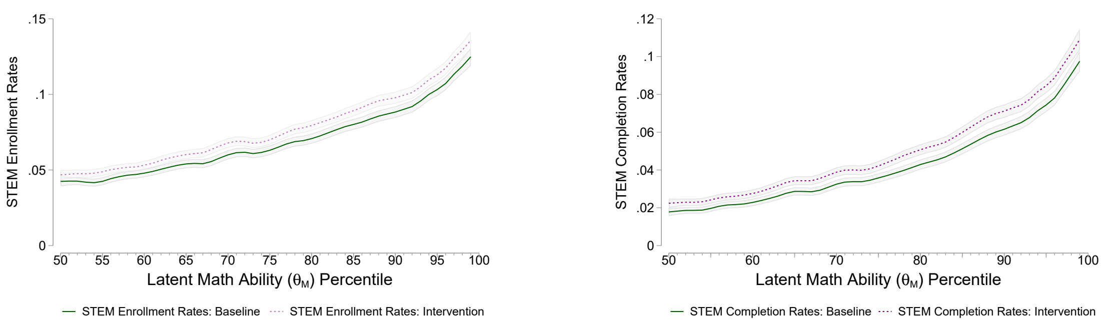
$$E[Y_i(p') - Y_i | D_{i,s}(p') = 0, D_{i,s} = 0] \times \underbrace{P[D_{i,s}(p') = 0, D_{i,s} = 0]}_{STEM\ Enrollment\ Never-Takers} \tag{19}$$

where $D_{i,s}$ is a dummy variable which equals one for student i enrolled in STEM. Equation (19) indicates that the aggregate effect of policy p' on outcome variable Y is given by the linear combination of the effect on STEM always-takers — students who enroll in STEM at baseline and under p', compliers, the students changing the enrollment decision due to the policy, and never-takers, who never enroll in STEM.[37] The effect of these interventions may further vary across the $\boldsymbol{\theta}$ distribution.

6.2 Impacts on STEM Participation

I first examine the effect of the self-efficacy intervention on women's STEM enrollment rates. The first panel of Figure 9 shows the share of high math ability women who would enroll in STEM upon an increase of 0.25 standard deviations in their self-efficacy. The simulated intervention would increase high θ_M women's STEM enrollment rates from 6.6% up to 7.3%, representing a 11% increase relative to baseline participation rates. Meanwhile, the self-efficacy increase would lead to increased STEM enrollment for women in the top math ability decile by one full percentage point, from 10.9% to 11.9%. Students would be primarily drawn away from majors in the humanities, social sciences and education into math-intensive STEM fields (Figure A.16).

Figure 9: Estimated Impacts of Intervention on STEM Participation



(a) Impacts on STEM Enrollment **(b)** Impacts on STEM Completion

Note: The first panel of Figure 9 presents the estimated impacts of the simulated self-efficacy intervention on STEM enrollment rates for women. The figure presents heterogeneous enrollment rates across different percentiles of the math ability distribution, both at baseline (green line) and under the policy intervention (purple line). The second panel presents the corresponding impacts of the simulated intervention on STEM completion rates. Shaded areas denote 95% confidence intervals.

Furthermore, the second panel of Figure 9 shows the simulated intervention would boost average STEM completion rates for women above the math ability median from 4% to 4.6%, or 15% of baseline completion rates. For women in the top math ability decile, the intervention would similarly boost graduation rates by a full percentage point, up to 9.1%. The overall effect on STEM completion rates emerges through two channels. First, 61.8% of 'compliers' would successfully

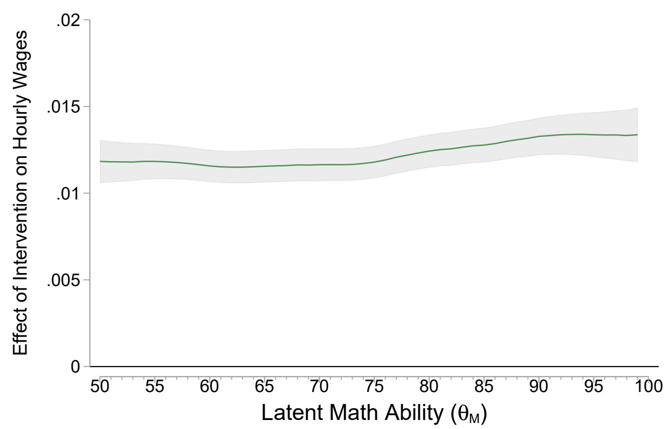
[37] Response types are given by students' initial major decisions. Compliers enroll in STEM as in policy p', but not under baseline policy p. Sorting-into-STEM patterns by math self-efficacy implies there are no defiers (Figure A.7).

complete a STEM degree. Moreover, among students originally enrolled in STEM — 'always-takers' — the simulated policy would increase their completion rates from 60% to 63.4% (Figure A.17), fitting in with the importance of self-efficacy in STEM completion documented earlier. I assess the robustness of these findings by estimating the effects of self-efficacy increases ranging from 0.05 to 0.5 standard deviations. Figure A.18 shows that larger-sized interventions having an positive linear impact on STEM participation rates.

6.3 Impacts on Labor Market Outcomes

In this context, a self-efficacy boost could affect labor market outcomes both through an increased likelihood of STEM completion but also through increased labor market productivity (Heckman et al., 2006). For instance, higher math self-efficacy could lead students in math-intensive jobs to improve their performance on math-related tasks and thus earn higher wages.

Figure 10: Estimated Impacts of Policy Intervention on Hourly Wages



Note: Figure 10 presents the estimated impacts of the simulated self-efficacy intervention on hourly wages for women above the median of the math ability distribution. Shaded areas denote 95% confidence intervals.

Altogether, the simulated intervention would boost average hourly wages by 1.2%, with largely homogeneous effects across the math ability distribution (Figure 10). The overall effect is comprised of heterogeneous impacts across the different response types (Figure A.19), yielding small positive effects for always-takers — driven through the productivity of self-efficacy for STEM enrollees — along with increasing impacts across the math ability distribution for policy compliers and a small positive impact on never-takers, explained by the wage returns to self-efficacy for non-STEM majors. These findings thus suggest that well-targeted skill development policies geared towards improving women's math self-efficacy could lead to increases in STEM participation while improving early-career labor market outcomes.[38]

[38] Using High School Longitudinal Study of 2009 data (Ingels et al., 2013), I show that self-efficacy is self-productive over time (Table A.9), suggesting that early-life interventions fostering math self-efficacy could have larger sustained effects on students' self-efficacy.

7 Conclusion

In recent years, women's under-representation in STEM has received increased attention in both the economics literature and in the policy discussion. In this paper, I have examined the interaction between multidimensional skills and college major choices, with the goal of understanding the factors driving women's participation in STEM majors. While previous work has focused on the importance of test scores and preferences as drivers of major choices, the importance of other skill dimensions has so far received scant attention in the existing literature.

To this end, I have introduced an empirical strategy which allows me to account for the fact that test scores measure latent skills with error and that skills are multiple in nature. I have thus been able to separately identify latent non-cognitive ability and mathematical self-efficacy. I have shown there is a relative shortfall of high-skilled women who are confident in their math abilities. This finding is relevant to the analysis of STEM participation, as students sort into these majors on both their math ability and their self-efficacy, as well. Furthermore, self-efficacy has a sizable effect in explaining female dropout from math-intensive fields, yet this pattern is largely muted for men. The shortfall of high-achieving women who are confident in their math skills reduces their participation in STEM majors. As such, future research should analyze the drivers of gender differences in mathematical self-efficacy given its importance in shaping gender gaps in STEM.

I have further shown that high math ability women would enjoy positive returns from STEM enrollment in the early career. Since math self-efficacy increases are associated with higher STEM participation rates for women, improving women's self-efficacy in mathematics could lead to improved early-career labor market outcomes. Further work on this front is thus needed to understand the types of interventions that can lead to sustained improvements in women's math self-efficacy.

References

Ahn, T., P. Arcidiacono, A. Hopson, and J. R. Thomas (2022). Equilibrium grade inflation with implications for female interest in STEM majors. Technical report, National Bureau of Economic Research.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

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. *Annual Review of Economics 4*(1), 185–223.

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

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2016). College attrition and the dynamics of information revelation. Technical report, National Bureau of Economic Research.

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

Aucejo, E. and J. James (2021). The path to college education: The role of math and verbal skills. *Journal of Political Economy 129*(10), 2905–2946.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review 84*(2), 191.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

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.

Berger, M. C. (1988). Predicted future earnings and choice of college major. *ILR Review 41*(3), 418–429.

Betz, N. E. and G. Hackett (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior 23*(3), 329–345.

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.

Blom, E., B. C. Cadena, and B. J. Keys (2021). Investment over the business cycle: Insights from college major choice. *Journal of Labor Economics 39*(4), 1043–1082.

Bordón, P., C. Canals, and A. Mizala (2020). The gender gap in college major choice in Chile. *Economics of Education Review 77*.

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.

Brenøe, A. A. and U. Zölitz (2020). Exposure to more female peers widens the gender gap in STEM participation. *Journal of Labor Economics 38*(4), 1009–1054.

Card, D. and A. A. Payne (2021). High school choices and the gender gap in stem. *Economic Inquiry 59*(1), 9–28.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). 2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review 44*(2), 361–422.

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.

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.

Delaney, J. M. and P. J. Devereux (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review 72*, 219–238.

Deming, D. J. and K. Noray (2020). Earnings dynamics, changing job skills, and STEM careers. *The Quarterly Journal of Economics 135*(4), 1965–2005.

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.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087.

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

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics 46*, 211–226.

Freyberger, J. (2018). Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies 85*(3), 1824–1851.

Goldin, C. (2015). Gender and the undergraduate economics major: Notes on the undergraduate economics major at a highly selective liberal arts college. *manuscript, April 12*.

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.

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.

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

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. Technical report, National Bureau of Economic Research.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

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., S. Urzua, and E. Vytlacil (2008). Instrumental variables in models with multiple outcomes: The general unordered case. *Annales d'Economie et de Statistique*, 151–174.

Horn, J. L. and J. J. McArdle (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research 18*(3), 117–144.

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

Humphries, J. E., J. S. Joensen, and G. F. Veramendi (2019). Complementarities in high school and college investments.

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

Ingels, S. J., D. J. Pratt, C. P. Alexander, D. M. Jewell, E. Lauff, T. L. Mattox, and D. Wilson (2014). Education Longitudinal Study of 2002 (ELS: 2002) Third follow-up data file documentation. NCES 2014-364. *National Center for Education Statistics*.

Ingels, S. J., D. J. Pratt, D. R. Herget, J. A. Dever, L. B. Fritch, R. Ottem, J. E. Rogers, S. Kitmitto, and S. Leinwand (2013). High School Longitudinal Study of 2009 (HSLS: 09) Base year to first follow-up data file documentation. appendixes. nces 2014-361. *National Center for Education Statistics*.

Jiang, X. (2021). Women in STEM: Ability, preference, and value. *Labour Economics 70*.

Joensen, J. S. and H. S. Nielsen (2016). Mathematics and gender: Heterogeneity in causes and consequences. *The Economic Journal 126*(593), 1129–1163.

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.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford publications.

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

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: are women really different from men? *Economics of Education Review 81*.

Lavy, V. and E. Sand (2018). On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics 167*, 263–279.

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

Mouganie, P. and Y. Wang (2020). High-performing peers and female STEM choices in school. *Journal of Labor Economics 38*(3), 805–841.

Mourifie, I., M. Henry, and R. Méango (2020). Sharp bounds and testability of a Roy model of STEM major choices. *Journal of Political Economy 128*(8), 3220–3283.

Olson, S. and D. G. Riordan (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. *Executive Office of the President*.

Oreopoulos, P. and K. G. Salvanes (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives 25*(1), 159–84.

Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology 21*(4), 325–344.

Patnaik, A., J. Venator, M. Wiswall, and B. Zafar (2020). The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice. *Journal of Econometrics*.

Patnaik, A., M. J. Wiswall, and B. Zafar (2020, August). College majors. Working Paper 27645, National Bureau of Economic Research.

Piatek, R. and P. Pinger (2016). Maintaining (locus of) control? Data combination for the identification and inference of factor structure models. *Journal of Applied Econometrics 31*(4), 734–755.

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.

Putnick, D. L. and M. H. Bornstein (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review 41*, 71–90.

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.

Rask, K. and J. Tiefenthaler (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review 27*(6), 676–687.

Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.

Rodríguez, J., F. Saltiel, and S. Urzúa (2022). Dynamic treatment effects of job training. *Journal of Applied Econometrics 37*(2), 242–269.

Rodríguez, J., S. Urzúa, and L. Reyes (2016). Heterogeneous economic returns to post-secondary degrees: Evidence from Chile. *Journal of Human Resources 51*(2), 416–460.

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

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2019). Ipums usa: Version 9.0 [dataset]. *Minneapolis, MN: IPUMS 10*, D010.

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. (2020). Gritting it out: The importance of non-cognitive skills in academic mismatch. *Economics of Education Review 78*.

Sheskin, D. J. (2020). *Handbook of parametric and nonparametric statistical procedures*. crc Press.

Shi, Y. (2018). The puzzle of missing female engineers: Academic preparation, ability beliefs, and preferences. *Economics of Education Review 64*, 129–143.

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 (2014). 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.

Webber, D. A. (2014). The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour Economics 28*, 14–23.

Williams, B. (2020). Identification of the linear factor model. *Econometric Reviews 39*(1), 92–109.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Wiswall, M. and B. Zafar (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

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

- SUPPLEMENTARY APPENDICES -

Appendix

A Additional Tables and Figures

Table A.1: Gender Differences in Prevalence of 'Almost Always' Responses to Math and English Self-Efficacy Questions, ELS 2002

Panel A: Full Sample

	Math		English	
	Male	Female	Male	Female
Question	(1)	(2)	(3)	(4)
Confident I can do an excellent job in tests	0.244	0.172***	0.238	0.266***
Confident I can understand most difficult material in texts	0.185	0.106***	0.183	0.188
Confident I can understand most complex material in class	0.218	0.147***	0.210	0.218
Confident I can do an excellent job on my assignments	0.245	0.198***	0.248	0.289***
Confident I can master the skills being taught in my class	0.258	0.204***	0.219	0.249***
Observations	5,404	5,987	5,188	5,763

Panel B: College Enrollees

	Math		English	
	Male	Female	Male	Female
Question	(1)	(2)	(3)	(4)
Confident I can do an excellent job in tests	0.356	0.231***	0.323	0.316
Confident I can understand most difficult material in texts	0.275	0.147***	0.236	0.228
Confident I can understand most complex material in class	0.322	0.187***	0.285	0.267
Confident I can do an excellent job on my assignments	0.367	0.264***	0.335	0.359
Confident I can master the skills being taught in my class	0.372	0.275***	0.297	0.309
Observations	1,904	2,545	1,830	2,468

Notes: Table A.1 presents the share of students in the ELS sample who respond 'almost always' to the questions included above. The first panel includes the full sample of students who answer these questions in the baseline survey, whereas the second panel restricts the analysis to those who eventually enroll in four-year college by age 20. In each panel, the first two columns present the responses to the math-related questions, and the last two columns present evidence for English-related questions. The stars in the even columns indicate the gender difference in the share who respond 'almost always' following from a two-sided t-test. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.2: Determinants of Observed Test Scores

	BY Math (1)	F1 Math (2)	GPA (3)	BY SE (4)	F1 SE (5)	Control (6)	Motivation (7)	Action (8)
Both Parents	0.000	0.011	0.001	-0.002	0.046**	0.006	0.001	-0.002
	(0.002)	(0.011)	(0.012)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Both Parents × Male	-0.000	0.019	0.013	0.000	-0.044	-0.016	0.000	0.005
	(0.003)	(0.016)	(0.018)	(0.032)	(0.029)	(0.032)	(0.032)	(0.033)
Parental Education	0.000	0.001	-0.001	0.004	-0.002	0.004	0.004	0.005*
	(0.000)	(0.001)	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Parental Education × Male	0.000	-0.001	-0.001	-0.006***	0.003	-0.005**	-0.006***	-0.006***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HH Income	0.001	0.009	-0.014	0.002	0.002	-0.003	0.005	0.008
	(0.001)	(0.009)	(0.010)	(0.017)	(0.016)	(0.018)	(0.017)	(0.018)
HH Income × Male	-0.002	-0.002	0.032**	0.058**	0.022	0.062**	0.058**	0.066**
	(0.002)	(0.013)	(0.016)	(0.029)	(0.025)	(0.029)	(0.029)	(0.030)
Constant	0.997***	0.940***	0.962***	0.761***	0.843***	0.750***	0.753***	0.732***
	(0.004)	(0.019)	(0.023)	(0.044)	(0.037)	(0.044)	(0.043)	(0.044)
Observations	4,599	4,599	4,599	4,599	4,599	4,599	4,599	4,599
R^2	0.001	0.004	0.002	0.007	0.003	0.006	0.006	0.007

Note: Table A.2 presents evidence from a linear regression on the drivers of having available test scores, observed self-efficacy and non-cognitive skill measures. Column headers refer to the skill measures used in the analysis. BY Math and F1 Math refer to the baseline and follow-up math exams, respectively. GPA indicates students' high school GPA. BY SE and F1 SE refer to students' baseline and follow-up math self-efficacy, respectively. Lastly, Control, Motivation and Action refer to students' action control, control expectation and their instrumental motivation in the baseline survey, respectively. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Sample Restrictions

Sample Restriction	Sample Size	Women	Men
Full Sample	16,197	8,107	8,090
Math Test Takers	16,146	8,084	8,062
Self-Efficacy Measures	13,466	6,885	6,581
Four-Year College Enrollment	5,188	2,903	2,285
Missing Educational Attainment	4,984	2,806	2,178
Missing Labor Market Outcomes	4,599	2,615	1,984

Note: Table A.3 shows the sample restrictions imposed on the baseline ELS sample result in the final sample used in the paper.

Table A.4: Determinants of Initial Major Choice: Multinomial Logit

Panel A. Determinants of Initial Major: Women

	STEM (1)	Life Sciences (2)	Business (3)	Health (4)
Both Parents	-0.008	0.018	-0.017	0.012
	(0.011)	(0.015)	(0.016)	(0.019)
Parental Education	0.004	0.003	-0.005	-0.006
	(0.002)	(0.002)	(0.003)	(0.003)
HH Income	0.004	-0.016	-0.003	-0.006
	(0.010)	(0.012)	(0.014)	(0.017)
Underrepresented Minority	0.027*	0.025	-0.017	-0.009
	(0.011)	(0.013)	(0.018)	(0.019)
Baseline Math Exam	0.016*	0.012*	-0.010	-0.026***
	(0.007)	(0.006)	(0.008)	(0.008)
HS GPA	0.000	0.014	-0.004	0.000
	(0.006)	(0.007)	(0.008)	(0.008)
Baseline Math Self-Efficacy	0.017***	0.008	0.028***	0.010
	(0.005)	(0.006)	(0.008)	(0.009)
Non-Cognitive Skills (PCA)	-0.003	0.013*	-0.018*	-0.001
	(0.005)	(0.007)	(0.008)	(0.010)

Panel B. Determinants of Initial Major: Men

	STEM (1)	Life Sciences (2)	Business (3)	Health (4)
Both Parents	-0.015	-0.011	0.038	0.018
	(0.024)	(0.015)	(0.026)	(0.013)
Parental Education	-0.002	0.007*	-0.002	-0.002
	(0.004)	(0.003)	(0.004)	(0.002)
HH Income	-0.005	-0.012	0.046*	-0.020*
	(0.022)	(0.013)	(0.021)	(0.010)
Underrepresented Minority	0.075**	0.025	0.008	-0.012
	(0.024)	(0.015)	(0.025)	(0.013)
Baseline Math Exam	0.043***	0.011	-0.015	-0.016***
	(0.011)	(0.006)	(0.009)	(0.005)
HS GPA	0.036***	0.022**	-0.010	0.012*
	(0.011)	(0.008)	(0.009)	(0.005)
Baseline Math Self-Efficacy	0.074***	-0.001	-0.011	-0.003
	(0.013)	(0.008)	(0.013)	(0.005)
Non-Cognitive Skills (PCA)	-0.034**	0.005	0.011	0.008
	(0.013)	(0.007)	(0.012)	(0.005)

Note: Table A.4 presents the estimated marginal effects from a multinomial logit examining the determinants of initial major choices for female (Panel A) and male (Panel B) students in the final ELS sample. The omitted category is enrollment in a major in the 'Other' category. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Determinants of Initial Major Choice: Multinomial Logit with Reading Test Score

Panel A. Determinants of Initial Major: Women

	STEM (1)	Life Sciences (2)	Business (3)	Health (4)
Both Parents	-0.008	0.018	-0.017	0.012
	(0.011)	(0.015)	(0.016)	(0.019)
Parental Education	0.004**	0.003	-0.004	-0.005
	(0.002)	(0.002)	(0.003)	(0.003)
HH Income	0.005	-0.016	0.001	-0.004
	(0.011)	(0.012)	(0.014)	(0.017)
Underrepresented Minority	0.027**	0.025*	-0.021	-0.011
	(0.011)	(0.013)	(0.018)	(0.019)
Baseline Math Exam	0.020***	0.013*	0.014	-0.010
	(0.007)	(0.008)	(0.009)	(0.010)
HS GPA	0.001	0.014*	-0.000	0.003
	(0.006)	(0.007)	(0.008)	(0.008)
Baseline Math Self-Efficacy	0.016***	0.008	0.021**	0.005
	(0.005)	(0.006)	(0.008)	(0.010)
Non-Cognitive Skills (PCA)	-0.002	0.014**	-0.014	0.001
	(0.005)	(0.007)	(0.008)	(0.010)
Baseline Reading Exam	-0.007	-0.003	-0.040***	-0.027***
	(0.006)	(0.008)	(0.009)	(0.009)

Panel B. Determinants of Initial Major: Men

	STEM (1)	Life Sciences (2)	Business (3)	Health (4)
Both Parents	-0.015	-0.011	0.038	0.018
	(0.024)	(0.015)	(0.026)	(0.013)
Parental Education	-0.000	0.007**	-0.001	-0.002
	(0.004)	(0.003)	(0.004)	(0.002)
HH Income	0.001	-0.013	0.048**	-0.020**
	(0.022)	(0.014)	(0.022)	(0.010)
Underrepresented Minority	0.073***	0.025	0.007	-0.012
	(0.024)	(0.015)	(0.025)	(0.013)
Baseline Math Exam	0.068***	0.011	0.003	-0.015***
	(0.013)	(0.007)	(0.012)	(0.006)
HS GPA	0.040***	0.022***	-0.007	0.012**
	(0.011)	(0.008)	(0.009)	(0.005)
Baseline Math Self-Efficacy	0.071***	-0.001	-0.016	-0.003
	(0.013)	(0.008)	(0.013)	(0.005)
Non-Cognitive Skills (PCA)	-0.032**	0.005	0.014	0.009*
	(0.013)	(0.007)	(0.012)	(0.005)
Baseline Reading Exam	-0.044***	0.001	-0.029***	-0.001
	(0.011)	(0.006)	(0.011)	(0.005)

Note: Table A.5 presents the estimated marginal effects from a multinomial logit examining the determinants of initial major choices for female (Panel A) and male (Panel B) students in the final ELS sample. It follows the results presented in Table A.4 while including baseline reading test scores as a control variable. The omitted category is enrollment in a major in the 'Other' category. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Estimated Parameters of the Latent Factor Distributions

Panel A. Factor Means			
	Male	Female	Joint
	(1)	(2)	(3)
θ_M	0.002	0.002	0.002
	(0.810)	(0.722)	(0.761)
		[0.854]	
θ_{SE}	0.414	0.003	0.180
	(0.946)	(0.983)	(0.988)
		[0.000]	
θ_{NC}	-0.037	-0.000	-0.016
	(0.438)	(0.591)	(0.530)
		[0.476]	

Panel B. Correlation Matrix: Women			
	θ_M	θ_{SE}	θ_{NC}
θ_M	1.000		
θ_{SE}	0.304	1.000	
θ_{NC}	0.135	0.444	1.000

Panel C. Correlation Matrix: Men			
	θ_M	θ_{SE}	θ_{NC}
θ_M	1.000		
θ_{SE}	0.488	1.000	
θ_{NC}	0.241	0.499	1.000

Note: The first panel of Table A.6 displays the mean, standard deviation for each latent factor for men (column 1) and women (column 2). The third column presents the corresponding mean and standard deviation for the distribution a latent factor that encompasses the underlying gender-specific distributions. The values in brackets present the p-values from a test of equality of the gender-specific factor means. Panels B and C present the correlation matrix for the latent factors for women and men, respectively.

Table A.7: Decomposition of Estimated Returns to STEM v. Alternative Majors by Gender

	Panel A. Women			
Estimate	Life Sciences (1)	Business (2)	Health (3)	Other (4)
Observed Differences	0.124	-0.031	-0.151	0.058
	[0.112,0.136]	[-0.041,-0.022]	[-0.160,-0.142]	[0.052,0.065]
Pairwise ATE	0.099	-0.089	-0.229	-0.006
	[0.091,0.107]	[-0.095,-0.083]	[-0.235,-0.223]	[-0.009,-0.003]
Selection Bias	0.013	0.071	0.093	0.061
	[0.002,0.025]	[0.063,0.079]	[0.084,0.102]	[0.054,0.067]
Sorting Gains	0.012	-0.014	-0.015	0.004
	[-0.004,0.027]	[-0.027,0.000]	[-0.029,-0.002]	[-0.008,0.016]
	Panel B. Men			
Estimate	Life Sciences (1)	Business (2)	Health (3)	Other (4)
Observed Differences	0.235	0.086	-0.054	0.255
	[0.226,0.245]	[0.080,0.092]	[-0.067,-0.041]	[0.250,0.259]
Pairwise ATE	0.264	0.088	-0.123	0.228
	[0.258,0.270]	[0.084,0.092]	[-0.132,-0.114]	[0.226,0.231]
Selection Bias	-0.037	-0.023	0.074	0.014
	[-0.048,-0.026]	[-0.029,-0.018]	[0.053,0.095]	[0.010,0.018]
Sorting Gains	0.008	0.022	-0.005	0.012
	[-0.001,0.017]	[0.015,0.029]	[-0.018,0.008]	[0.006,0.019]

Note: Table A.7 presents the decomposition introduced in Heckman et al. (2018) of observed differences across STEM majors and all other alternative majors into pairwise average treatment effects — which capture differences in potential wages for individuals enrolled in either STEM or in respective major k —, selection bias and sorting gains. All parameters follow from the estimated model. The first panel presents evidence for women and the second panel presents results for men. 95% confidence intervals are presented in brackets.

Table A.8: Distribution of First- and Second-Best Choices (in %) by Gender

	Panel A. Women				
First Best	STEM	Life Sciences	Business	Health	Other
Second Best	(1)	(2)	(3)	(4)	(5)
STEM	—	10.2	8.6	7.9	13.7
Life Sciences	13.7	—	11.9	12.0	20.0
Business	16.9	16.4	—	19.0	29.8
Health	18.3	19.4	22.0	—	36.4
Other	51.0	54.0	57.5	61.1	—
Total	100	100	100	100	100
	Panel B. Men				
First Best	STEM	Life Sciences	Business	Health	Other
Second Best	(1)	(2)	(3)	(4)	(5)
STEM	—	25.8	24.1	20.0	34.5
Life Sciences	13.9	—	11.1	10.3	16.4
Business	24.5	21.1	—	22.0	37.6
Health	7.2	6.6	7.7	—	11.5
Other	54.5	46.5	57.1	47.6	—
Total	100	100	100	100	100

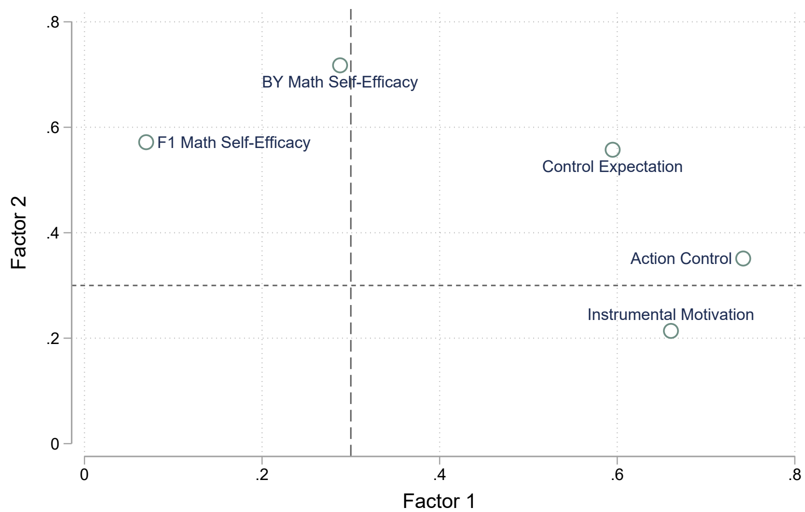
Note: For individuals enrolled in different majors (presented in columns (1)-(5)), Table A.8 documents the share of students with different second best options. The first panel presents evidence for women enrolled across the five major groupings in the model (STEM, life sciences, business, health and 'Other') and the second panel shows evidence for men.

Table A.9: The Process of Math Self-Efficacy Development

	Full Sample	Males	Females
	(1)	(2)	(3)
Math Test Score	0.120***	0.142***	0.095***
	(0.010)	(0.014)	(0.015)
Math Self-Efficacy	0.206***	0.223***	0.190***
	(0.012)	(0.016)	(0.017)
Math Identity	0.132***	0.133***	0.131***
	(0.012)	(0.016)	(0.017)
Math Interest	0.037**	0.034*	0.041*
	(0.011)	(0.015)	(0.017)
Math Utility	0.017	0.014	0.016
	(0.010)	(0.013)	(0.014)
Background Characteristics	Yes	Yes	Yes
R^2	0.1657	0.1841	0.1286
Observations	12,645	6,219	6,426

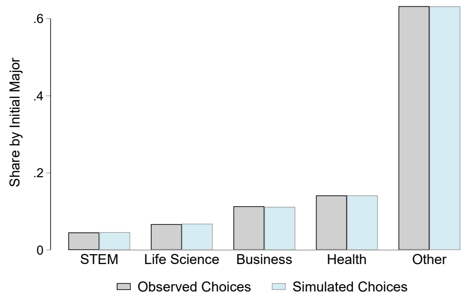
Source: High School Longitudinal Study of 2009 (HSLS). Note: Table A.9 uses HSLS data from the baseline and first-follow up survey rounds to examine how baseline characteristics (math self-efficacy, identity, interest, utility and exam performance) influence students' math self-efficacy at follow-up in the following equation: $SE_{i,t+1} = \alpha_1 SE_{i,t} + \alpha_2 ID_{i,t} + \alpha_3 INT_{i,t} + \alpha_4 U_{i,t} + \alpha_5 T_{i,t} + \eta X_{i,t} + v_{i,t}$. All variables are standardized within the sample. The first column presents evidence for the full sample. The second and third columns focus on the male and female samples, respectively. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: Exploratory Factor Analysis: Non-Cognitive Skill Measures

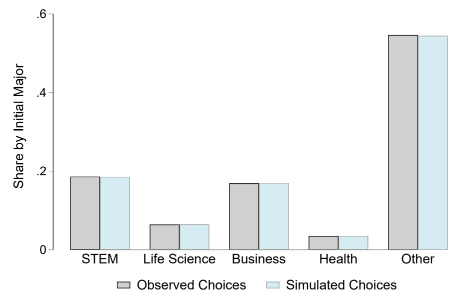


Note: Figure A.1 presents the estimated loadings from exploratory factor analysis of the three non-cognitive skill measures available in the ELS along with the baseline and follow-up math self-efficacy measures with orthogonal factors. The solid horizontal and dashed vertical lines are placed at 0.30, as loadings with an absolute value above 0.30 are considered significant (Sheskin, 2020).

Figure A.2: Goodness of Fit: Initial Majors



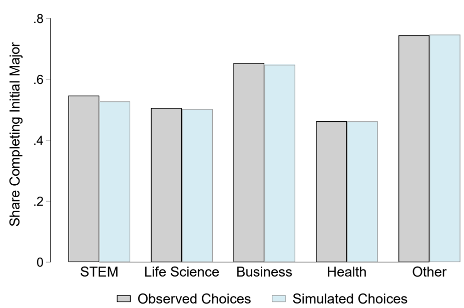
(a) Initial Major Choices: Women



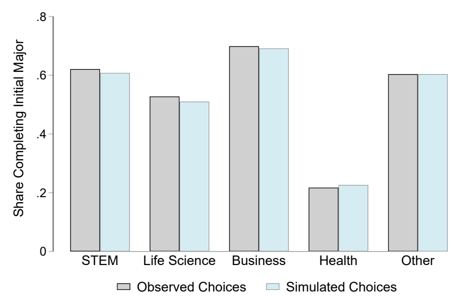
(b) Initial Major Choices: Men

Note: Figure A.2 compares the observed share of students initially enrolled in different majors to the model-based estimated shares of initial major enrollment for females (Panel (a)) and males (Panel (b)).

Figure A.3: Goodness of Fit: Initial Major Completion Rates by Gender



(a) % Completing Initial Major: Women

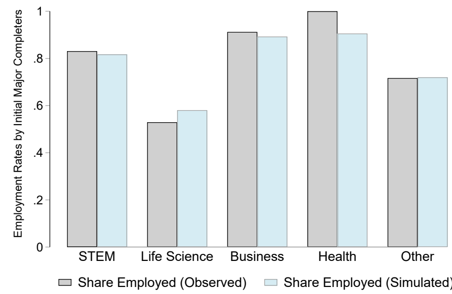


(b) % Completing Initial Major: Men

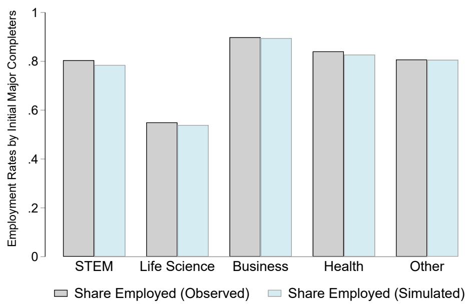
Note: The first panel of Figure A.3 compares the observed share of students initially enrolled in different majors who end up completing those majors to the model-based estimated shares of initial-major completion for females (Panel (a)) and males (Panel (b)).

Figure A.4: Goodness of Fit: Employment Rates Among Initial Major Completers by Household Income

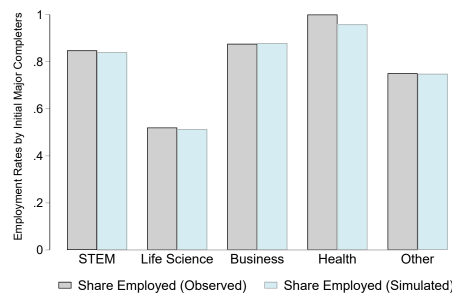
(a) % Employed by Major: Low-HH Income Women



(b) % Employed by Major: Low-HH Income Men



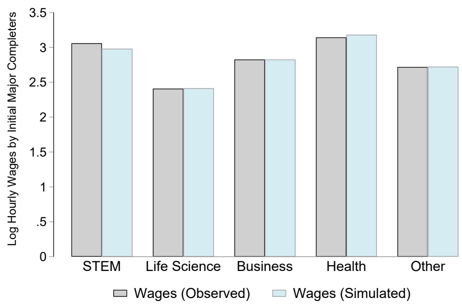
(c) % Employed by Major: High-HH Income Women



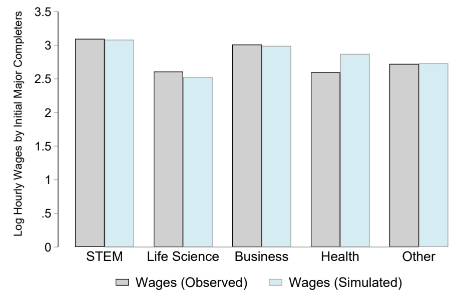
(d) % Employed by Major: High-HH Income Men

Note: Figure A.4 presents the observed proportion of students who are employed in 2011 among those who completed their initial major by whether their household income in the baseline survey was above or below the US household income median in 2001. These proportions are compared to the corresponding model-based simulated share of initial-major completers who are employed. The first two panels present evidence for below-median households whereas the last two panels present corresponding evidence for above-median households.

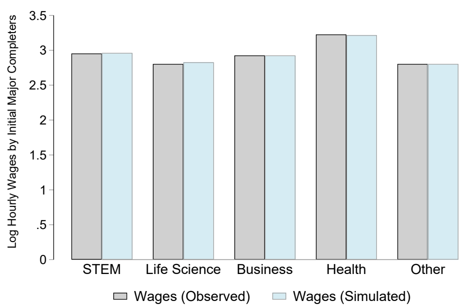
Figure A.5: Goodness of Fit: Hourly Wages Among Initial Major Completers



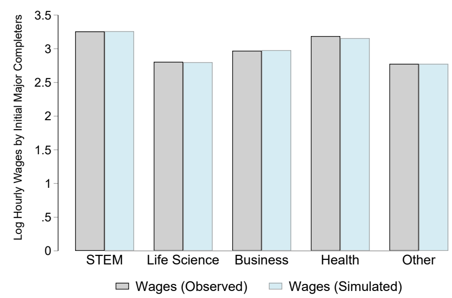
(a) Wages by Major: Low-HH Income Women



(b) Wages by Major: Low-HH Income Men



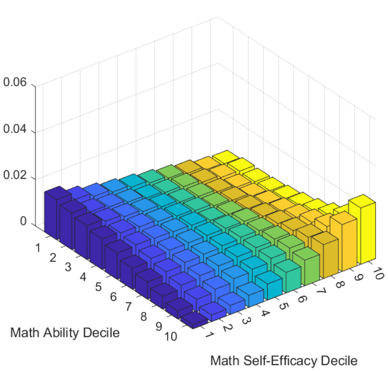
(c) Wages by Major: High-HH Income Women



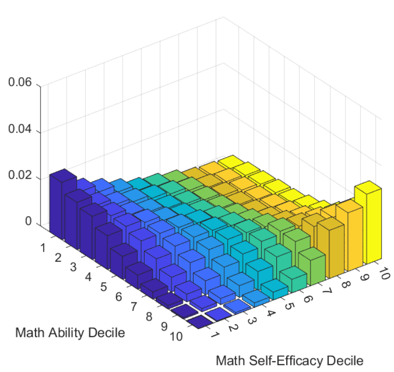
(d) Wages by Major: High-HH Income Men

Note: Figure A.5 presents average hourly wages for employed individuals who completed their initial major by whether their household income in the baseline survey was above or below the US household income median in 2001. These proportions are compared to the corresponding model-based average wages for initial major completers who are employed. The first two panels present evidence for below-median households whereas the last two panels present corresponding evidence for above-median households.

Figure A.6: Gender-Specific Distribution of Latent Math Ability and Self-Efficacy



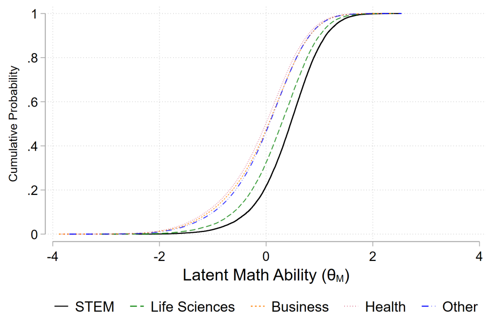
(a) Joint Distribution of θ_M and θ_{SE}: Women



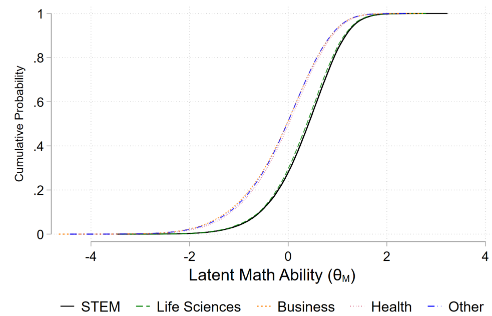
(b) Joint Distribution of θ_M and θ_{SE}: Men

Note: Figure A.6 presents the joint density of the gender-specific math ability and self-efficacy by gender, documenting the share of individuals pertaining to the gender-specific joint decile of the two latent factor distributions (first two columns of

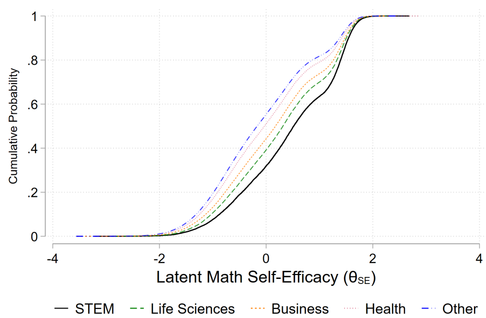
Figure A.7: Initial Major Choices by Gender



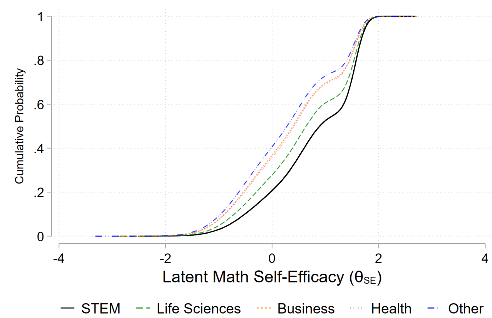
(a) θ_M by Major: Women



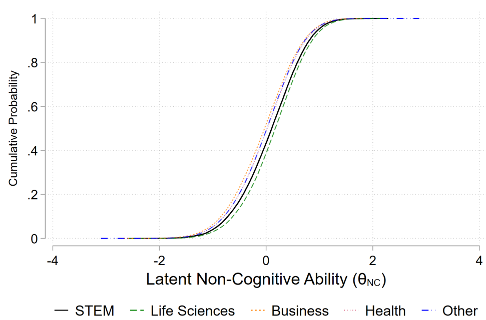
(b) θ_M by Major: Men



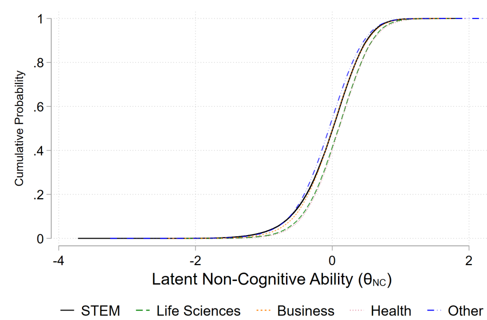
(c) θ_{SE} by Major: Women



(d) θ_{SE} by Major: Men



(e) θ_{NC} by Major: Women

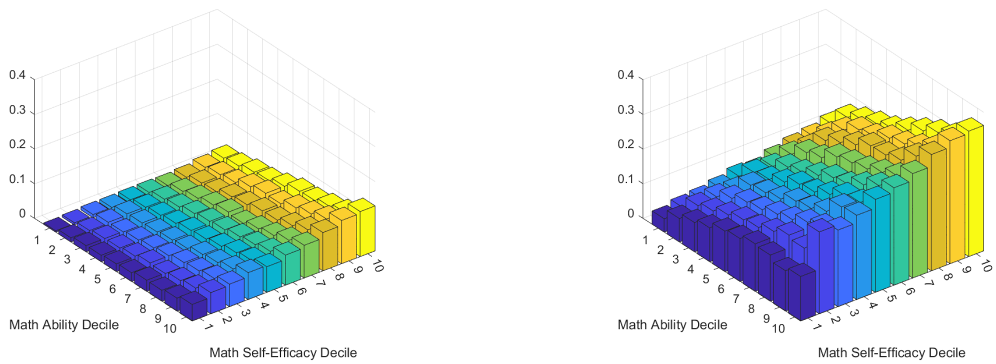


(f) θ_{NC} by Major: Men

Panel A in Table A.6). The first panel presents results for women and the second panel presents evidence for men.

Note: Figure A.7 presents the cumulative distribution function of the latent factors ($\boldsymbol{\theta}$) for female and male students enrolled in different initial majors. The first two panels show evidence for latent math ability, panels (c) and (d) present evidence for mathematical self-efficacy and the last two panels show the CDF for latent non-cognitive skills.

Figure A.8: STEM Enrollment Rates by Gender-Specific θ_M and θ_{SE} Deciles

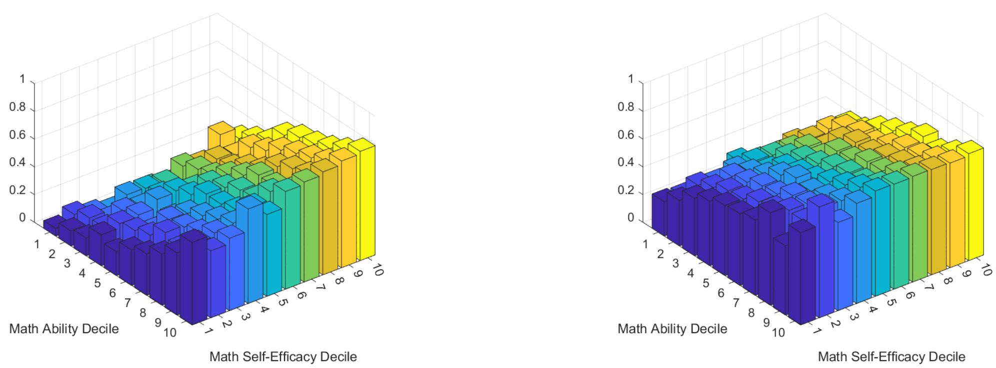


(a) STEM Enrollment Rates: Women

(b) STEM Enrollment Rates: Men

Note: Figure A.8 shows the share of women and men initially enrolled in a math-intensive major at each decile of the joint gender-specific θ_M and θ_{SE} distribution (first two columns of Panel A in Table A.6). The first panel presents results for women and the second panel presents evidence for men.

Figure A.9: STEM Completion Rates Among Initial Enrollees: by Gender

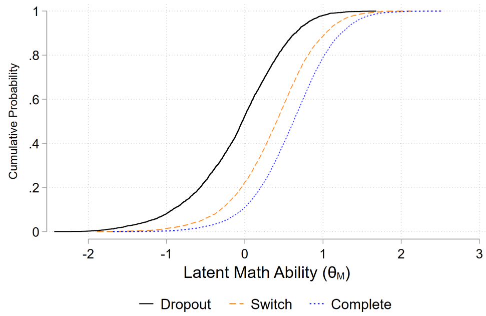


(a) STEM Completion Rates: Women

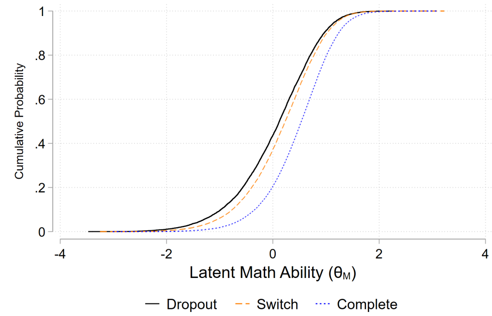
(b) STEM Completion Rates: Men

Note: Figure A.9 shows the share of women and men who complete a STEM degree after initially enrolling in a math-intensive major. The share of conditional completers is presented across each decile of the joint gender-specific math ability and self-efficacy distribution (first two columns of Panel A in Table A.6). The first panel presents results for women and the second panel presents evidence for men.

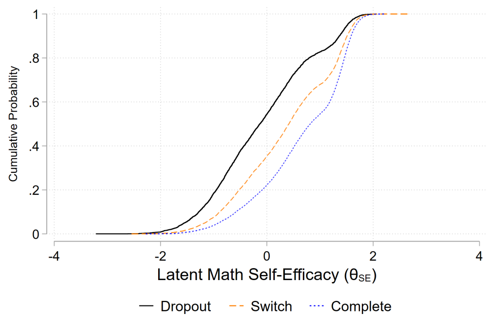
Figure A.10: STEM Completion Outcomes Among Initial STEM Enrollees by Gender



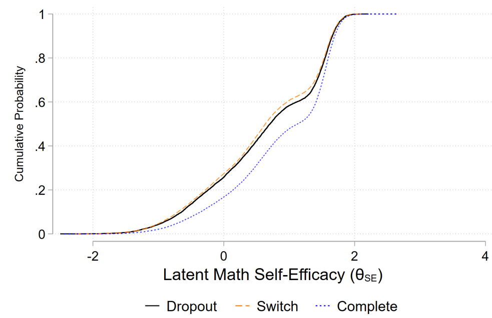
(a) θ_M by Outcome: Women



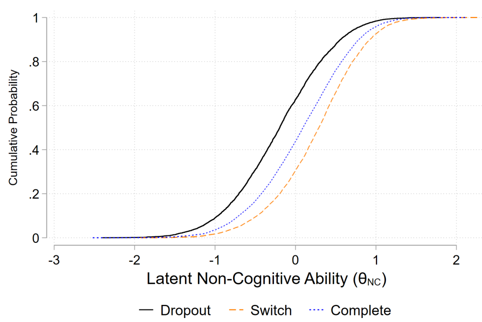
(b) θ_M by Outcome: Men



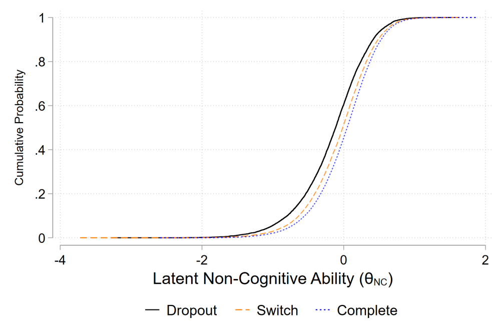
(c) θ_{SE} by Outcome: Women



(d) θ_{SE} by Outcome: Men



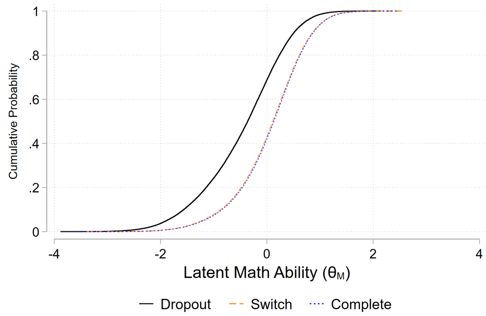
(e) θ_{NC} by Outcome: Women



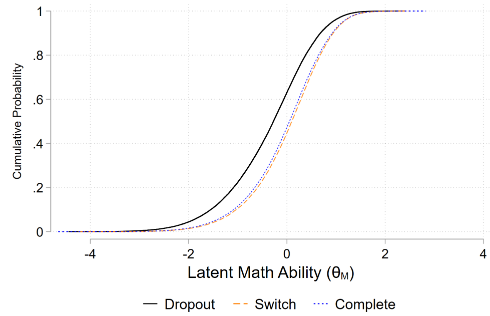
(f) θ_{NC} by Outcome: Men

Note: Figure A.10 presents the cumulative distribution function of the latent factors ($\boldsymbol{\theta}$) for female and male students enrolled in STEM depending on whether they complete a STEM major, switch to a different major or drop out of college altogether. The first two panels show evidence for latent math ability, panels (c) and (d) present evidence for mathematical self-efficacy and the last two panels show the CDF for latent non-cognitive skills.

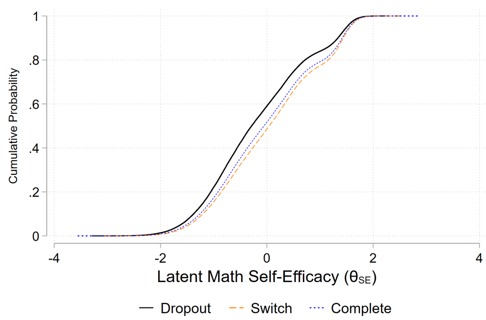
Figure A.11: Graduation Outcomes Among Initial Non-STEM Enrollees by Gender



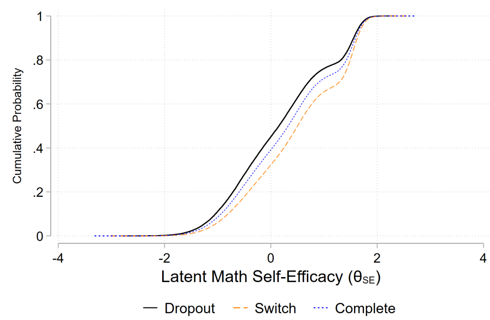
(a) θ_M by Outcome: Women



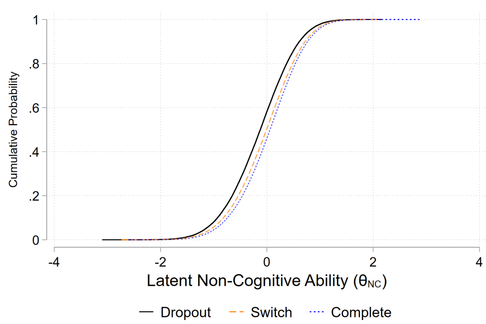
(b) θ_M by Outcome: Men



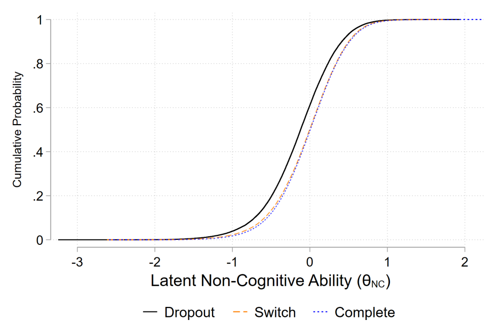
(c) θ_{SE} by Outcome: Women



(d) θ_{SE} by Outcome: Men



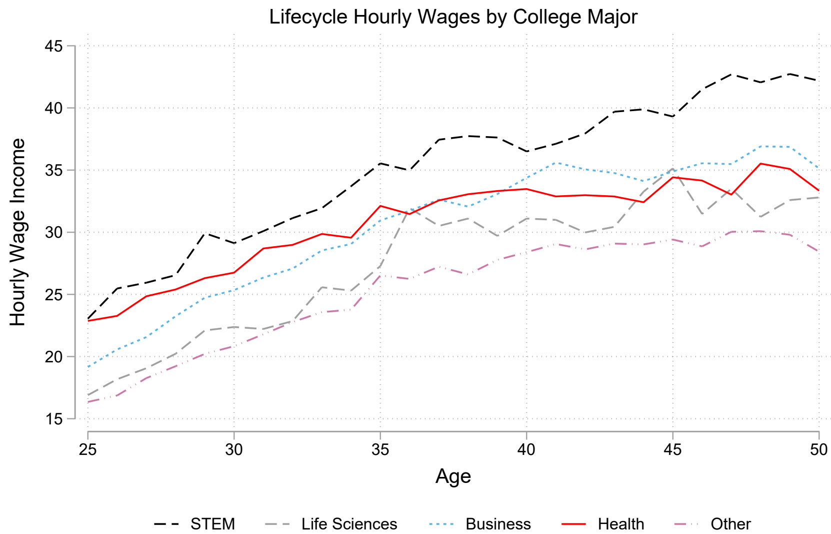
(e) θ_{NC} by Outcome: Women



(f) θ_{NC} by Outcome: Men

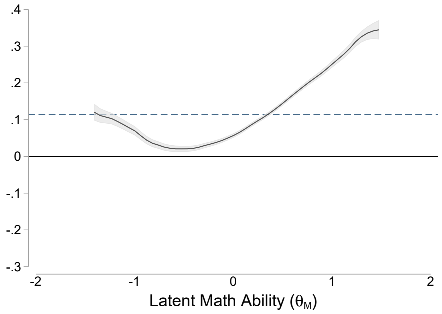
Note: Figure A.11 presents the cumulative distribution function of the latent factors ($\boldsymbol{\theta}$) for female and male students enrolled in non-STEM majors depending on whether they complete their initial major, switch to a different major or drop out of college altogether. The first two panels show evidence for latent math ability, panels (c) and (d) present evidence for mathematical self-efficacy and the last two panels show the CDF for latent non-cognitive skills.

Figure A.12: Hourly Wages by College Major and Age, American Community Survey Data

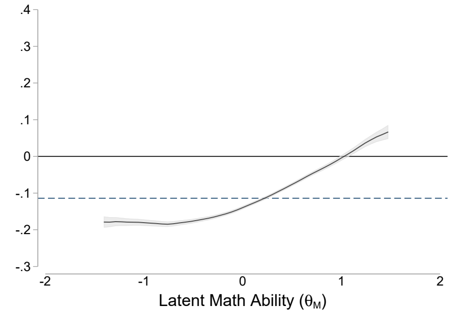


Source: American Community Survey, 2011-2012 (Ruggles et al., 2019).
Note: Figure A.12 presents average hourly wages by college major for college graduates aged 25-50. College majors are aggregated into five categories, encompassing STEM, Life Sciences, Business, Health and the remaining majors.

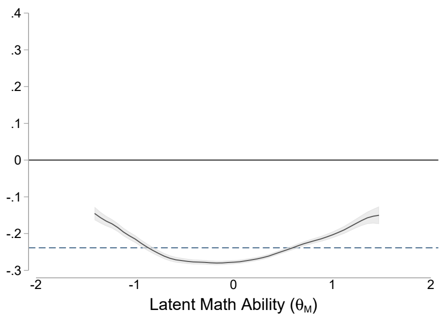
Figure A.13: Heterogeneous Returns to STEM Enrollment Across θ_M Distribution for Women



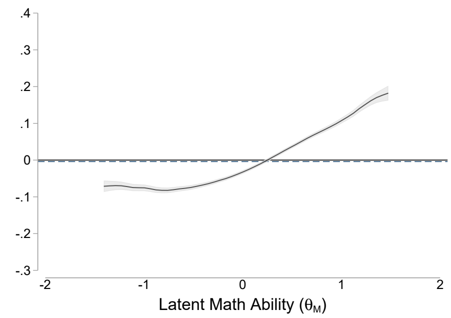
(a) v. Life Sciences



(b) v. Business



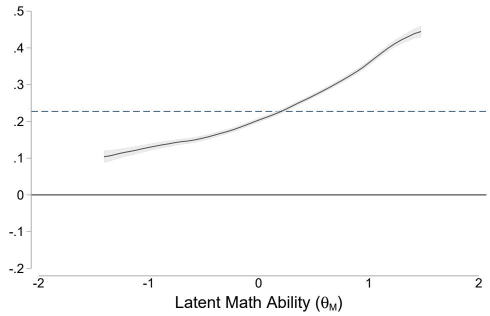
(c) v. Health



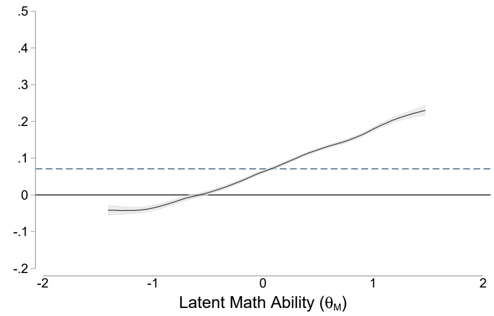
(d) v. Other

Note: Figure A.13 presents heterogeneous returns to STEM enrollment relative to life sciences, business, health and 'Other' majors for women across the latent math ability (θ_M) distribution. Dashed lines in each panel denote the corresponding average treatment effect parameter. Shaded areas denote 95% confidence intervals.

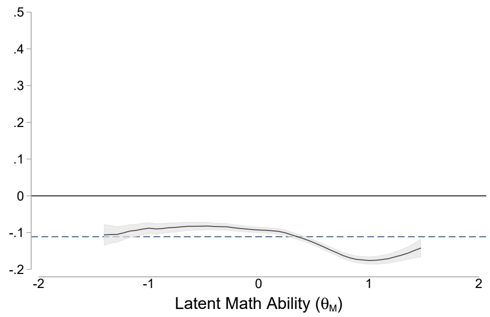
Figure A.14: Heterogeneous Returns to STEM Enrollment Across θ_M Distribution for Men



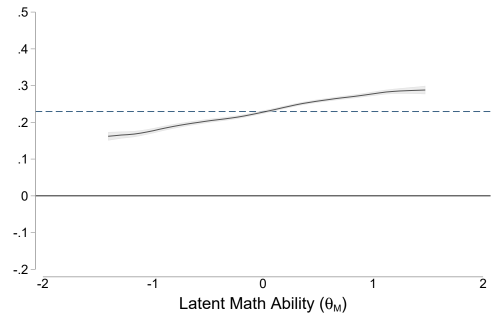
(a) v. Life Sciences



(b) v. Business



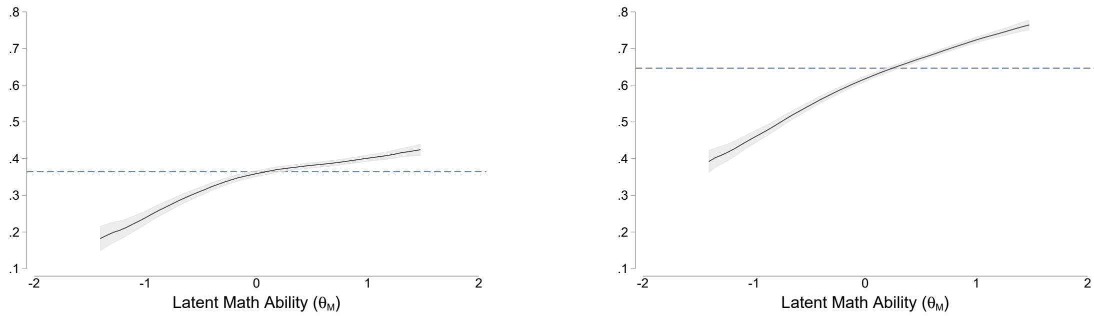
(c) v. Health



(d) v. Other

Note: Figure A.14 presents heterogeneous returns to STEM enrollment relative to life sciences, business, health and 'Other' majors for men across the latent math ability (θ_M) distribution. Dashed lines in each panel denote the corresponding average treatment effect parameter. Shaded areas denote 95% confidence intervals.

Figure A.15: Heterogeneous Returns to STEM Completion for Male STEM Enrollees

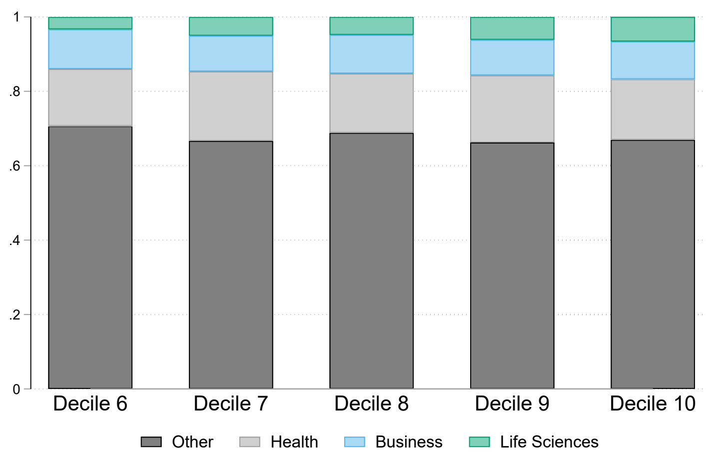


(a) Returns v. Switching

(b) Returns v. Dropout

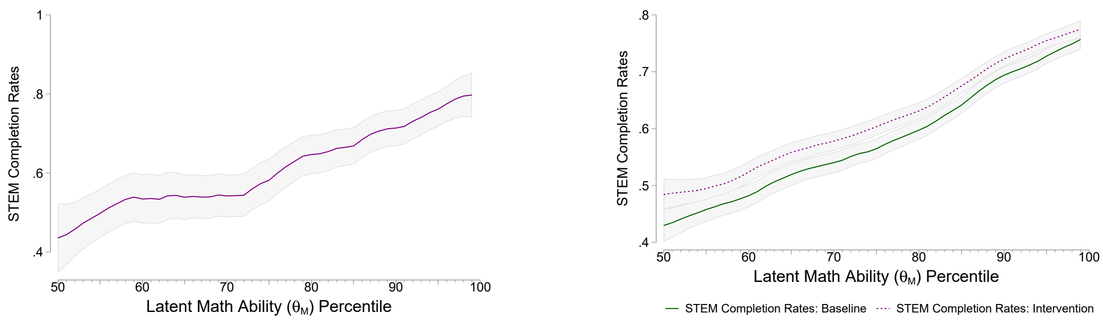
Note: The first panel of Figure A.15 presents heterogeneous returns to STEM completion relative to switching to a different major (equation (17)) for men. The second panel presents corresponding evidence relative to dropping out of college altogether (equation (18). Dashed lines denote the corresponding average treatment effect in each panel. Shaded areas denote 95% confidence intervals

Figure A.16: Complier Types Across Initial Majors



Note: Figure A.16 presents how the estimated self-efficacy intervention would affect the share of compliers across different deciles of the math ability distribution who would be switching out of different majors.

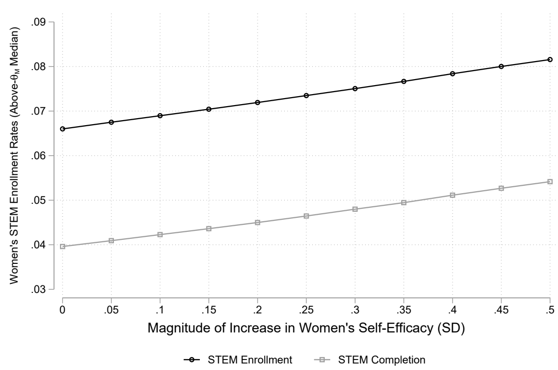
Figure A.17: Estimated Impacts of Policy Intervention on STEM Completion Rates



(a) Impact on STEM Completion: Compliers **(b)** Impact on STEM Completion: Always-Takers

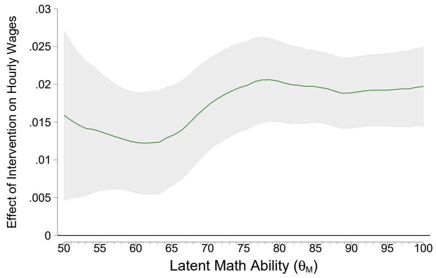
Note: The first panel of Figure A.17 presents graduation rates for 'compliers' only under the simulated intervention, as their STEM completion rates would have equaled zero in absence of the policy. The second panel presents the estimated impacts of the simulated self-efficacy intervention on STEM completion rates for policy 'always-takers', as defined in Section 6. Shaded areas denote 95% confidence intervals.

Figure A.18: Estimated Impacts of Self-Efficacy Interventions on STEM Participation

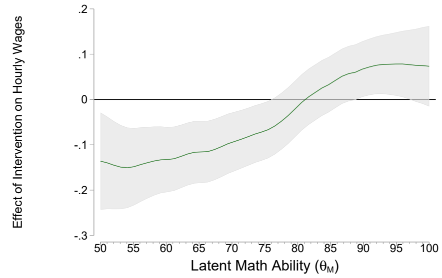


Note: Figure A.18 presents the estimated impacts of self-efficacy interventions which would increase women's mathematical self-efficacy from 0.05 to 0.5 standard deviations on their STEM enrollment and completion rates.

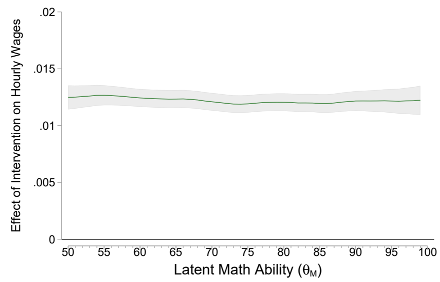
Figure A.19: Impacts of Intervention on Hourly Wages by Response Type



(a) Impact on Hourly Wages: Always-Takers



(b) Impact on Hourly Wages: Compliers



(c) Impact on Hourly Wages: Never-Takers

Note: Figure A.19 presents the estimated impacts of the simulated self-efficacy intervention on hourly wages for policy 'always-takers' (Panel A), compliers (Panel B) and 'never-takers' (Panel C), as defined in Section 6. Shaded areas denote 95% confidence intervals.

B Initial College Enrollment by Gender

B.1 Reduced Form Evidence

To examine whether students differentially select into four-year college enrollment based on their baseline skills by gender, I take advantage of data from the ELS:2002 survey. In particular, I impose the same sample restrictions as in Section 2.1, yet instead of restricting the sample to four-year college enrollees by the second follow-up survey (age 20), I restrict the sample to students who had completed a high school degree by age 20. As such, the sample size becomes larger, reaching 9,627 individuals, where 4,599 students belong to the main sample (i.e. enrolled in four year college), 1,870 were enrolled in two-year college at age 20, and 3,158 had not yet enrolled in college by age 20. In this sample, a larger share of women (49.7%) than men (45.4%) have enrolled in four-year college by age 20, fitting in with the existing evidence in the literature.

To examine whether female college enrollees are more positively selected into college on their skill measures than their male peers, I estimate the following regression:

$$Test_i = \alpha + \beta Male_i + \gamma College_i + \lambda Male_i \times College_i + v_i \tag{B.1}$$

where β measures the extent to which test scores differ by gender in the overall sample, γ captures the extent to which four-year college enrollees are positively selected based on their test scores, and λ recovers whether such selection patterns differ by gender. I estimate equation (B.1) separately for the eight measures included in the measurement system in the paper, encompassing math test scores, GPA, math self-efficacy and non-cognitive skill measures.

I present the results in Table B.1. In the full sample, men earn higher math test scores and have higher math self-efficacy than women, whereas the opposite is true for GPA and in the action control measure. Moreover, students are positively selected into four-year college enrollment across these eight measures, with larger differences emerging in math test scores and GPA than in the non-cognitive skill measures and math self-efficacy. Importantly, the estimated coefficient on the four-year sample and male interaction is small in magnitude (below 0.08 standard deviations across all measures), and only significant for GPA. These results thus indicate that while students are positively selected into four-year college enrollment, the extent to which sorting-into-college patterns by skill level differ by gender are largely muted. In light of these results, the findings presented in the paper hold for both male and female four-year college enrollees by age 20.

Table B.1: Sorting into Four-Year College by Skills and Gender, ELS

	BY Math (1)	F1 Math (2)	BY SE (3)	F1 SE (4)	GPA (5)	Control (6)	Motivation (7)	Action (8)
Male	0.565***	0.026	0.002	0.007	0.069*	0.006	-0.012	0.011
	(0.207)	(0.026)	(0.017)	(0.013)	(0.041)	(0.017)	(0.017)	(0.017)
College Enrollee	0.942***	0.958***	0.361***	0.320***	0.915***	0.403***	0.277***	0.345***
	(0.023)	(0.023)	(0.024)	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)
College Enrollee × Male	0.003	0.027	0.012	0.039	0.083**	0.003	0.044	-0.010
	(0.036)	(0.035)	(0.035)	(0.037)	(0.034)	(0.035)	(0.036)	(0.035)
N	9627	9627	9627	9627	9627	9627	9627	9627

Note: Table B.1 presents evidence on the extent to which students sort into college enrollment by age 20 based on their skills (equation (B.1)). The sample is restricted to ELS 2002 respondents who had completed a high school degree by age 20, encompassing 9,627 individuals. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Model of College Enrollment and Major Choices

To assess whether the main findings presented in the paper are robust to the inclusion of non-college-enrollees, I estimate a version of the model introduced in Section 3 which directly accounts for students' college enrollment choices. In this model, I posit that a vector of latent abilities, encompassing students' math ability, non-cognitive skills and their math self-efficacy, affects their educational choices and associated labor market outcomes. The alternative version of the framework also models educational choices sequentially, where upon completing high school, students first decide whether to enroll in four-year college, enroll in two-year college or not enroll in higher education altogether (Heckman et al., 2006; Urzua, 2008; Rodríguez et al., 2016). Let $V_{i,h}^{G}$ be the utility for student i of gender G from enrolling in higher-education level $h \in \mathcal{H}$, where $V_{i,h}^{G}$ is given by:

$$V_{i,h} = \beta_h X_{i,h} + \alpha_h \boldsymbol{\theta_i} + \varepsilon_{i,h} \quad for \ \ h \in \mathcal{H} \tag{B.2}$$

where $X_{i,h}$ encompasses household and individual characteristics affecting the higher education choice, $\boldsymbol{\theta_i}$ captures the vector of latent ability, and ε_i^h is an error term that is independent of $\{X_{i,h}; \boldsymbol{\theta_i}\}$ and across the three higher education options. Conditional on $\{X_{i,h}\boldsymbol{\theta_i}\}$, initial enrollment choices are unordered and students thus enroll in the initial level with the highest utility: $D_{i,h} = argmax_{h \in \mathcal{H}} V_{i,h}$.

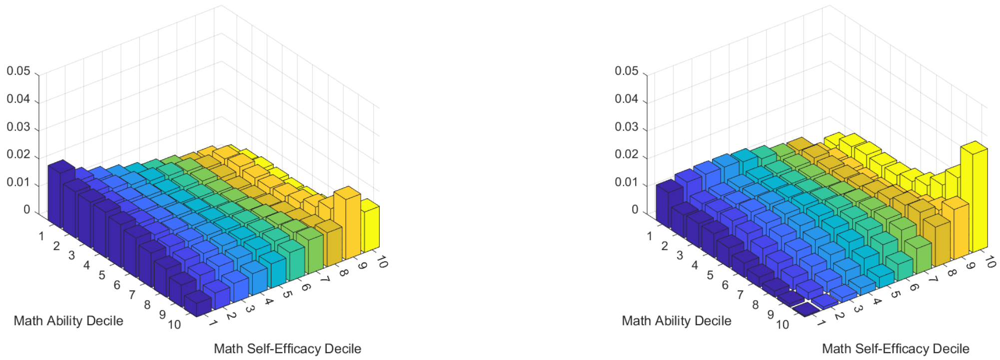
The rest of the model proceeds as the one introduced in Section 3. First, four-year college enrollees select an initial college major.[39,40] These students then decide whether to continue in four-year college or dropout and college continuers lastly choose whether to remain in their initial major or to switch to a different degree. Upon finishing their educational choices, students enter the labor market, make an employment decision and earn hourly wages. The alternative version of the model directly follows the structure of educational choices and labor market outcomes introduced in equations (1)-(6). I also posit a measurement system in which observed skill measures are a linear outcome of students' latent abilities ($\boldsymbol{\theta_i}$) and of their background characteristics, directly following the structure introduced in equations (7)-(9) in the main draft. The identification arguments introduced in Section 3 still hold in this context and I allow for the mean of the latent factors for male to differ from zero. The model implementation and estimation approach follow from the model presented in Section 3, where I simulate a sample of 200 from the posterior distribution of estimated model parameters.

Model Results. I use the re-estimated model parameters and replicate the main results presented in Section 4. First, in Figure B.1, I present the joint distribution of math ability and self-efficacy. Fitting in with the results in the main draft, I find a positive correlation between these two latent skill components, yet with a stronger correlation for men. Among students in the top math-ability decile, a larger share of men are in the top decile of mathematical self-efficacy (38.4%) compared to the corresponding share for women (13.3%).

[39] Two-year college enrollees' completion outcomes are not directly modeled, as their educational choices are not the focus of the paper. These students thus enroll in two-year college and I then allow them examine their labor market outcomes. The same structure holds for students who do not enroll in college by age 20.

[40] I remark that I additionally restrict the analysis of major choices to three options, encompassing math-intensive STEM majors, business and other majors in order to reduce the number of parameters to be estimated.

Figure B.1: Joint Distribution of Latent Math Ability and Self-Efficacy by Gender

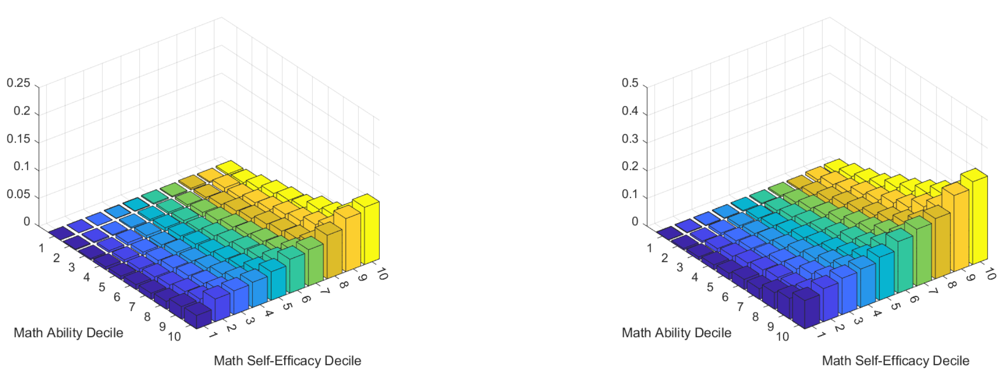


(a) Joint Distribution of θ_M and θ_{SE}: Women **(b)** Joint Distribution of θ_M and θ_{SE}: Men

Note: Figure B.1 presents the joint distribution of math ability and self-efficacy. The math ability and self-efficacy deciles are given by the joint distribution encompassing men and women. The figure documents the share of women (Panel A) and men (Panel B) pertaining to each joint decile of the two latent factor distributions.

Next, I assess whether the STEM enrollment patterns documented in Figure 5 remain similar when incorporating non-college enrollees as part of the analysis. Figure B.2 presents the share of students who initially enroll in STEM majors at four-year colleges across the baseline distribution of their math ability and their self-efficacy. First, I remark that the overall share of STEM enrollment is lower than in the main model, as this version of the model incorporates the option of initially starting in a two-year college or not enrolling. These differences are particularly pronounced among lower-skilled students, who are more likely to not enroll in four-year college altogether. Nonetheless, I still find strong evidence that students in the top joint decile of the math ability and self-efficacy distribution are far more likely to have enrolled in STEM than their lower-skilled peers across both skill dimensions. In fact, these patterns hold for male and female students, fitting in with the results presented in Figure 5.

Figure B.2: STEM Enrollment Rates by θ_M and θ_{SE} by Gender

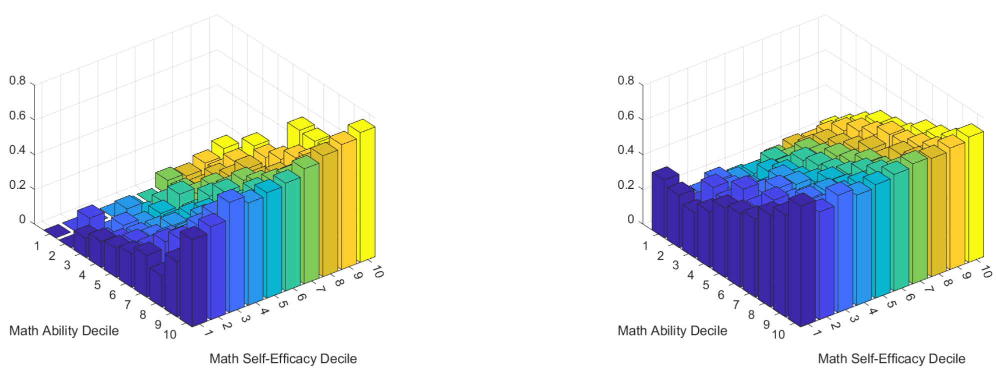


(a) STEM Enrollment Rates: Women **(b)** STEM Enrollment Rates: Men

Note: Figure B.2 shows the share of women and men who initially enroll a STEM degree. The share of STEM enrollees is presented across each decile of the joint math ability and self-efficacy distribution, where the deciles are defined by the joint distribution of ability encompassing both men and women.

In Figure B.3, I analyze how math ability and self-efficacy jointly affect the likelihood of STEM completion for students who started in these majors. Similar to the main model, I still find strong evidence that higher-math ability students are more likely to successfully complete a STEM degree, a finding that holds for both men and women. At the same time, I find that math self-efficacy strongly determines the likelihood of successful STEM completion for women, whereas this skill dimension is far less consequential for men. As such, since math self-efficacy plays a far stronger role in driving women's exit from STEM, and this is not the case for men, the main results regarding STEM-sorting patterns based on skills remain similar across the two versions of the model. As such, the results presented in Figures B.1-B.3 show that the main findings in Section 4 hold in an alternative model which incorporates students' choice of initial enrollment levels.

Figure B.3: STEM Completion Rates Among Initial Enrollees: by Gender



(a) STEM Completion Rates: Women **(b)** STEM Completion Rates: Men

Note: Figure B.3 shows the share of women and men who complete a STEM degree after initially enrolling in a math-intensive major. The share of conditional completers is presented across each decile of the joint math ability and self-efficacy distribution, where the deciles are defined by the joint distribution of ability encompassing both men and women. The first panel presents results for women. The second panel presents evidence for men.

C 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); Heckman et al. (2006); Williams (2020). In the measurement system presented in equations (7)-(9), 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 $\boldsymbol{X}$ and the gender superscript implicit.

First, note that the covariance between the two observed self-efficacy measures is given by:

$$Cov(SE_1, SE_2) = \alpha_{SE,1}\alpha_{SE,2}\sigma^2_{\theta_{SE}} \tag{C.1}$$

where $\sigma^2_{\theta_{SE}}$ represents the variance of the latent math self-efficacy factor. Since there are three unknown right-hand side parameters to be identified, the covariance between the self-efficacy measures does not suffice for identification. As such, I further rely on the six observed covariances between the two self-efficacy measures and three available non-cognitive skill measures, which are given by:

$$Cov(SE_j, NC_k) = \alpha_{SE,j}\alpha_{SE,k}\sigma^2_{\theta_{SE}} + \alpha_{SE,j}\gamma_{NC,k}\sigma_{\theta_{SE},\theta_{NC}} \tag{C.2}$$

where $\sigma_{\theta_{SE},\theta_{NC}}$ captures the covariance of the latent math self-efficacy and non-cognitive skill factor. Equations (C.1)-(C.2) indicate there are seven observed covariances across the self-efficacy and non-cognitive skill factors, yet ten parameters need to be identified (eight factor loadings, the variance of θ_{SE} and the covariance of θ_{SE} and θ_{NC}). Since latent factors have no scale of their own, Carneiro et al. (2003) note that one of the factor loadings can be normalized to unity to set the scale of each latent factor ($\alpha_{SE,1} = 1$ and $\gamma_{NC,1} = 1$).[41] As discussed in Section 3.2, I follow the reduced form evidence presented in Figure A.1 and additionally assume that instrumental motivation is a dedicated measures of latent non-cognitive ability ($\alpha^{SE}_{NC2} = 0$). As such, the remaining seven parameters can be identified from the observed covariances. Next, note that the covariance between observed non-cognitive measures has the following structure:

$$Cov(NC_k, NC_{k'}) = \alpha_{SE,k}\alpha_{SE,k'}\sigma^2_{\theta_{SE}} + \gamma_{NC,k}\gamma_{NC,k'}\sigma^2_{\theta_{NC}} + (\alpha_{SE,k}\gamma_{NC,k'} + \alpha_{SE,k'}\gamma_{NC,k})\sigma_{\theta_{SE},\theta_{NC}} \tag{C.3}$$

where $\sigma^2_{\theta_{NC}}$ captures the covariance of the latent non-cognitive ability factor (θ_{NC}). Since all other right-hand side parameters are identified from the covariances introduced in equations (C.1)-(C.2), $\sigma^2_{\theta_{NC}}$ is identified from any covariance between the observed non-cognitive skill measures.

As shown in equation (9), the three math measures (including the two test scores and high school GPA) load on all three factors. As such, nine factor loadings need to be identified along with the variance of the math ability factor and two remaining covariances between θ_M and θ_{SE} as well as the covariance between θ_M and θ_{NC}. First, note that the covariance across the math measures is given by:

$$\begin{aligned} Cov(M_l, M_{l'}) &= \alpha_{SE,l}\alpha_{SE,l'}\sigma^2_{\theta_{SE}} + \gamma_{NC,l}\gamma_{NC,l'}\sigma^2_{\theta_{NC}} + \eta_{M,l}\eta_{M,l'}\sigma^2_{\theta_M} + (\alpha_{SE,l}\gamma_{NC,l'} + \alpha_{SE,l'}\gamma_{NC,l})\sigma_{\theta_{SE},\theta_{NC}} \\ &\quad + (\alpha_{SE,l}\eta_{M,l'} + \alpha_{SE,l'}\eta_{M,l})\sigma_{\theta_{SE},\theta_M} + (\gamma_{NC,l}\eta_{M,l'} + \gamma_{NC,l'}\eta_{M,l})\sigma_{\theta_{NC},\theta_M} \end{aligned} \tag{C.4}$$

where $\sigma_{\theta_{SE},\theta_M}$ denotes the covariance between θ_{SE} and θ_M and $\sigma_{\theta_{NC},\theta_M}$ captures the covariance between θ_{NC} and θ_M. Since the three covariances across observed math test scores do not suffice for identifying the remaining unknown parameters, I further take advantage of the covariance of

[41] I normalize the loadings for the baseline self-efficacy measure and for control expectation, respectively.

math test scores and non-cognitive skill measures, which are given by:

$$
\begin{aligned}
Cov(M_l, NC_k) \quad = \quad & \alpha_{SE,l}\alpha_{SE,k}\sigma^2_{\theta_{SE}} + \gamma_{NC,l}\gamma_{NC,k}\sigma^2_{\theta_{NC}} + (\alpha_{SE,l}\gamma_{NC,k} + \alpha_{SE,k}\gamma_{NC,l})\sigma_{\theta_{SE},\theta_{NC}} \\
& +(\alpha_{SE,k}\eta_{M,l})\sigma_{\theta_{SE},\theta_M} + (\gamma_{NC,k}\eta_{M,l})\sigma_{\theta_{NC},\theta_M} \qquad \text{(C.5)}
\end{aligned}
$$

As such, the twelve covariances defined in equations (C.4)-(C.5) allow me to recover the remaining factor loadings (I normalize $\eta_{M,1} = 1$ for the baseline math test score), the variance of θ_M ($\sigma^2_{\theta_M}$) and the two remaining covariances of the latent factors.[42] Upon securing identification of the factor loadings, along with the variance and covariance of the three latent factors, I can take advantage of the variance of each observed measure to identify the variance of the corresponding error terms ($\sigma^2_{SE_j}, \sigma^2_{NC_k}, \sigma^2_{M_l}$). Upon identifying all the factor loadings and the variance of each component of latent ability, I rely on the identification arguments presented in Freyberger (2018) to non-parametrically identify the distribution of the latent factors and error terms.[43]

The correlation between unobserved abilities is identified through the following linear relationships between the latent factors:

$$\theta_{NC} = \delta\theta_{SE} + \theta_V \qquad \text{(C.6)}$$

$$\theta_M = \phi\theta_{SE} + \theta_V + \theta_U \qquad \text{(C.7)}$$

where (θ_V, θ_U) are additional factors assumed to be independent of $(\theta_{SE}, \theta_{NC}, \theta_M)$. Note that each non-cognitive skill measure can be re-written as:

$$
\begin{aligned}
NC_k &= \alpha_{SE,k}\theta_{SE} + \gamma_{NC,k}\theta_{NC} + e_{NC,k} \\
NC_k = \alpha_{SE,k}\theta_{SE} + \gamma_{NC,k}(\delta\theta_{SE} + \theta_V) + e_{NC,k} &= \phi_{SE,k}\theta_{SE} + \gamma_{NC,k}\theta_V + e_{NC,k}
\end{aligned}
$$

where $\phi_{SE,k} = \alpha_{SE,k} + \gamma_{NC,k}\delta$, and $\alpha_{SE,k}$ and $\gamma_{NC,k}$ are identified through the arguments presented above. The three equations for the observed non-cognitive skill measures have four unknowns (δ and $\alpha_{SE,k}$ for the three measures). As discussed in Section 3, δ is identified by assuming that instrumental motivation is a dedicated measure of non-cognitive skills ($\alpha_{SE,k} = 0$). Similarly, each math test score can be re-written as:

$$
\begin{aligned}
M_l &= \alpha_{SE,l}\theta_{SE} + \gamma_{NC,l}\theta_{NC} + \eta_{M,l}\theta_M + e_{M,l} \\
M_l &= \alpha_{SE,l}\theta_{SE} + \gamma_{NC,l}(\delta\theta_{SE} + \theta_V) + \eta_{M,l}(\phi\theta_{SE} + \theta_V + \theta_U) + e_{M,l} \\
M_l &= \xi_{SE,l}\theta_{SE} + (\gamma_{NC,l} + \eta_{M,l})\theta_V + \eta_{M,l}\theta_U + e_{M,l}
\end{aligned}
$$

where $\xi_{SE,l} = \alpha_{SE,l} + \gamma_{NC,l}\delta + \eta_{M,l}\phi$. Following the arguments outlined above, there are three math measures and four unknowns (η and the three $\alpha_{SE,l}$ loadings); η is identified by upon assuming GPA does not directly depend on latent math self-efficacy ($\alpha_{SE,l} = 0$). The model is estimated using the three orthogonal factors, $\theta_{SE}, \theta_V, \theta_U$, and the distribution of the correlated factors is recovered through the convolution of the these three factors.

[42] As discussed in Section 3, I additionally normalize the self-efficacy loading of high school GPA to zero ($\eta_{M,3} = 0$).

[43] Freyberger (2018) extends the identification arguments in Kotlarski (1967) to a context with correlated factors.

D Gender Gaps in Latent Abilities

This appendix presents the formal argument behind the identification of gender differences in the latent ability vector, which allows for a direct comparison of the estimated latent factors across males and females. In particular, consider the self-efficacy measure j score for males m and females f:

$$SE_j^m = \varphi_{SE,j}^m + \alpha_{SE,j}^m \theta_{SE}^m + e_{SE,j}^m \tag{D.1}$$
$$SE_j^f = \varphi_{SE,j}^f + \alpha_{SE,j}^f \theta_{SE}^f + e_{SE,j}^f \tag{D.2}$$

where the dependence on observables for simplicity is omitted for simplicity and the mean of the error terms is equal to zero. Let μ_{SE}^m and μ_{SE}^f denote the means of latent self-efficacy for men and women, respectively, with $\Delta_{SE} = \mu_{SE}^m - \mu_{SE}^f$ capturing their difference. As discussed in Section 3, identifying Δ_{SE} requires a scalar invariance assumption, which in this case amounts to assuming $\varphi_{SE,j}^m = \varphi_{SE,j}^f$. As such,

$$E(SE_{i,j}^m) - E(SE_{i,j}^f) = (\varphi_{SE,j}^m - \varphi_{SE,j}^f) + \alpha_{SE,j}^m \Delta_{SE} \tag{D.3}$$

Since the left-hand side in (D.3) can be computed from the observed self-efficacy measures and the factor loadings ($\alpha_{SE,j}^m$) are identified through the arguments in Appendix C, assuming that $\mu_{SE}^f = 0$ (or any other normalization) allows me to directly identify Δ_{SE}. The same intuition follows for the identification of the differences in the means of latent non-cognitive skills (θ_{NC}). In this case, for non-cognitive skill measure k:

$$NC_k^m = \varphi_{NC,k}^m + \alpha_{SE,k}^m \theta_{SE}^m + \gamma_{NC,k}^m \theta_{NC}^m + e_{NC,k}^m \tag{D.4}$$
$$NC_k^f = \varphi_{NC,k}^f + \alpha_{SE,k}^f \theta_{SE}^f + \gamma_{NC,k}^f \theta_{NC}^f + e_{NC,k}^f \tag{D.5}$$

Letting μ_{NC}^m and μ_{NC}^f capture the means of non-cognitive skills for men and women, respectively, the difference is similarly given by: $\Delta_{NC} = \mu_{NC}^m - \mu_{NC}^f$. Assuming scalar invariance and normalizing μ_{NC}^f to equal zero implies that the gender gap in non-cognitive measure k is given by:

$$E(NC_k^m) - E(NC_k^f) = \alpha_{SE,j}^m \Delta_{SE} + \gamma_{NC,k}^m \Delta_{NC} \tag{D.6}$$

where Δ_{SE} is identified in equation (D.3), the factor loadings are identified through the arguments in Appendix C and the left-hand side can be computed directly in the data, allowing for the identification of Δ_{NC}. Lastly, the same arguments follow for the identification of differences in latent math ablity (θ_M). For math test score l:

$$M_l^m = \varphi_{M,l}^m + \alpha_{SE,l}^m \theta_{SE}^m + \gamma_{NC,l}^m \theta_{NC}^m + \eta_{M,l}^m \theta_M^m + e_{M,l}^m \tag{D.7}$$
$$M_l^f = \varphi_{M,l}^f + \alpha_{SE,l}^f \theta_{SE}^f + \gamma_{NC,l}^f \theta_{NC}^f + \eta_{M,l}^f \theta_M^f + e_{M,l}^f \tag{D.8}$$

with μ_M^m and μ_M^f denoting the means of latent math ability by gender and the difference being given by: $\Delta_M = \mu_M^m - \mu_M^f$. The measurement invariance assumption and the normalization of the mean of women's latent math ability to equal zero implies that for math test score m the gender gap is given by:

$$E(M_l^m) - E(M_l^f) = \alpha_{SE,l}^m \Delta_{SE} + \gamma_{NC,l}^m \Delta_{NC} + \eta_{SE,l}^m \Delta_M \tag{D.9}$$

where the left-hand side is computed in the data. The gender gap in latent self-efficacy and non-cognitive skills are identified in equations (D.3)-(D.6), and the factor loadings are already identified (Appendix C), implying that Δ_M is identified as well. The empirical implementation of (D.1)-(D.9) first assumes the factor means for both genders are equal to zero, and the assumption is then relaxed along with the imposition of scalar invariance to recover the average gender gap for each latent factor. The identification of the gender differences in the latent abilities requires having access to one observed measure for each skill dimension. Nonetheless, since I observe multiple measures for each skill construct, I follow the logic outlined in this section and compute the gap in latent skills by averaging across all measures that correspond to a latent factor.

E Details on MCMC Estimator

In this section, I describe the MCMC algorithm used for model estimation (Hansen et al., 2004; Heckman et al., 2006). For simplicity, I introduce the estimator in the context of a standard Roy model, encompassing the option to pursue a STEM major (S) or a non-STEM major (S), with one latent factor θ. The model is given as follows:

$$
\begin{aligned}
I &= Z\gamma + C, \\
Y_S &= X_0\beta_{S,0} + X_1\beta_{S,1} + \varepsilon_S, \\
Y_N &= X_0\beta_{N,0} + X_1\beta_{N,1} + \varepsilon_N. \\
D &= \mathbb{1}[I > 0]
\end{aligned}
$$

where I captures the net value of pursuing a STEM major, where $I > 0$ indicates STEM participation ($\{D = 1\} \equiv S$) and $I \leq 0$ denotes non-STEM participation ($\{D = 0\} \equiv N$).

Assume a factor structure of the form:

$$
\begin{aligned}
C &= \theta\alpha_C + U_C \\
\varepsilon_S &= \theta\alpha_S + U_S \\
\varepsilon_N &= \theta\alpha_N + U_N
\end{aligned}
$$

where $U_S \perp\!\!\!\perp U_N \perp\!\!\!\perp U_C$, and $\theta \perp\!\!\!\perp (U_N, U_S, U_C)$, and

$$
\begin{aligned}
&U_S \sim N(0, \sigma^2_{U_S}), \\
&U_N \sim N(0, \sigma^2_{U_N}), \\
&U_C \sim N(0, \sigma^2_{U_C}).
\end{aligned}
$$

As a measurement system, consider the test score equation:

$$T = W\eta + \theta + U_{T_1}$$

where $U_{T_1} \sim N(0, \sigma^2_{U_{T_1}})$.

Thus, the likelihood function can be written as:

$$
\begin{aligned}
f(Y_j, T_j, D_j; \Theta) &= \int f(Y_j, T_j, D_j|\theta_j)dF(\theta_j) \\
&= \int f(Y_j, D_j|\theta_j)f(T_j|\theta_j)dF(\theta_j).
\end{aligned}
$$

Under the assumptions outlined above:

$$
\begin{aligned}
\Gamma(Y, T, D; \Theta) &= \prod_{j=1}^{N} \int \left[\frac{1}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{1}{2\sigma_T^2}(T_j - W_j\eta - \theta_j)^2\right)\right] \\
&\quad \left[\frac{1}{\sqrt{2\pi}\sigma_{u_S}} \exp\left(-\frac{1}{2\sigma^2_{u_S}}(Y_{S,j} - X_j\beta_S - \alpha_S\theta_j)^2\right)\Phi\left(-Z_j\gamma - \alpha_C\theta_j\right)\right]^{D_j}
\end{aligned}
$$

$$\left[\frac{1}{\sqrt{2\pi}\sigma_{u_N}}\exp\left(-\frac{1}{2\sigma_{u_N}^2}(Y_{0,j}-X_j\beta_N-\alpha_N\theta_j)^2\right)\left(1-\Phi\left(-Z_j\gamma-\alpha_C\theta_j\right)\right)\right]^{1-D_j}dF(\theta_j).$$

The block structure associated with the likelihood function is thus:

$$\begin{aligned}
f(\alpha,\beta,\tau,\gamma,\eta,\theta/Y,T,D) &\propto f(Y,T,D|\alpha,\beta,\tau,\gamma,\eta,\theta)f(\alpha,\beta,\tau,\gamma,\eta,\theta)\\
&\propto f(Y,D|\alpha,\beta,\tau,\gamma,\eta,\theta)f(T|\alpha,\beta,\tau,\gamma,\eta,\theta)f(\alpha,\beta,\tau,\gamma,\eta,\theta)\\
&\propto f(Y|D,\alpha,\beta,\tau,\gamma,\eta,\theta)f(D|\alpha,\beta,\tau,\gamma,\eta,\theta)f(T|\alpha,\beta,\tau,\gamma,\eta,\theta)f(\alpha,\beta,\tau,\gamma,\eta,\theta)\\
&\propto f(Y|D,\alpha,\beta,\tau,\gamma,\eta,\theta)f(D|\alpha,\beta,\tau,\gamma,\eta,\theta)f(T|\alpha,\beta,\tau,\gamma,\eta,\theta)f(\alpha,\beta,\tau,\gamma,\eta,\theta)\\
&= \tau_{u_S}^{n_{(D_i=1)}/2}\exp\left(-\tau_{u_S}\sum_{(D_i=1)}(Y_i^S-X_i\beta_S-\alpha_S\theta_i)^2\right)\\
&\quad\times\tau_{u_N}^{n_{(D_i=0)}/2}\exp\left(-\tau_{u_N}\sum_{(D_i=0)}(Y_i^N-X_i\beta_N-\alpha_N\theta_i)^2\right)\\
&\quad\times\prod_{i=1}^{n}\left[\Phi\left(-Z_i\gamma-\alpha_v\theta_i\right)\right]^{D_i}\left[\left(1-\Phi\left(-Z_i\gamma-\alpha_v\theta_i\right)\right)\right]^{1-D_i}\\
&\quad\times\tau_{u_{T1}}^{n/2}\exp\left(-\tau_{u_{T1}}(\sum(T_i-W_i\eta-\theta_i)^2\right)\\
&\quad\times\tau_f^{n/2}\exp(-\tau_f(\sum\theta_i^2)\exp(-\tau_{u_S})\exp(-\tau_{u_N})\exp(-\tau_{u_{T1}})\\
&\quad\times\left(\frac{1}{10}\right)^{1/2}\exp\left\{-\frac{1}{10}(\alpha_N)^2\right\}\left(\frac{1}{10}\right)^{1/2}\exp\left\{-\frac{1}{10}(\alpha_S)^2\right\}\\
&\quad\left(\frac{1}{10}\right)^{1/2}\exp\left\{-\frac{1}{10}(\alpha_C)^2\right\},
\end{aligned}$$

where I explicitly impose a set of prior distributions. Using the block structure, the formula for the conditional posteriors is given by:

1. Outcome equations:

$$f(\beta_i/\alpha_i,\tau_i,\theta,Y,D)\propto\exp\left\{-\tau_i\left(\sum_{j:D=i}(Y_j-X_j\beta_i-\theta_j)^2\right)\right\}\quad\text{for } i=S,N$$

so

$$\beta_i/.\sim N\left(\frac{\sum_{j:D=i}x_j(y_j-\alpha_i\theta_j)}{\sum_{j:D=i}x_jx_j'},\left(\tau_i\sum_{j:D=i}x_jx_j'\right)^{-1}\right),\tag{E.1}$$

$$f(\alpha_i/\beta_i,\tau_i,f,Y_i)\propto\exp\left\{-\frac{1}{2}\tau_i\sum_{j:D=i}(y_j-x_j\beta_i-\alpha_i\theta_j)^2-\frac{1}{2}\frac{1}{10}\alpha_i^2\right\}\quad\text{for } i=S,N.$$

Let $\widetilde{Y}=Y-X\beta$, thus

$$\exp\left\{-\frac{1}{2}\tau_{u_i}\sum_{j:D=i}(\widetilde{y}_j-\alpha_i\theta_j)^2-\frac{1}{2}\frac{1}{10}\alpha_i^2\right\}$$

Thus

$$\alpha_i/\beta_i, \tau_i, \theta, Y, D \sim N(\widehat{\alpha}_i, \overline{\Sigma}_i) \tag{E.2}$$

where $\overline{\Sigma}_i = \left(\tau_i \sum_{j:D=i} \theta_j^2 + \frac{1}{10}\right)^{-1}$ and $\widehat{\alpha}_i = \overline{\Sigma}_i \left(\tau_i \sum_{j:D=i} \theta_j \widetilde{y}_j\right)$.

2. Measurement system:

$$f(\eta/\tau_T, \theta, W, T) \propto \exp\left(-\tau_T \sum_{j=1}^{N} (T_j - W_j\eta - \theta_j)^2\right)$$

then

$$\eta/. \sim N(\left(W'W\right)^{-1}(W'(T-\theta)), \tau_{u_{T1}}^{-1}(W'W)^{-1}) \tag{E.3}$$

3. Decision model:

Let D^* be the latent variable. The completion for D. is thus defined as:

$$\begin{aligned} f(\gamma, \alpha_C, D^*/D) &\propto f(D/\alpha_C, \gamma, D^*) f(D^*/\alpha_C, \gamma) f(\gamma) f(\alpha_C) \\ &= \left(\frac{1}{10}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{1}{10}(\alpha_C)^2\right\} \prod_{j=1}^{N} \left[1(D_j^* > 0)\phi\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right)\right]^{D_j} \\ &\qquad \left[1(D_j^* < 0)\phi\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right)\right]^{1-D_j} \end{aligned}$$

Then,

$$\begin{aligned} f(\gamma/\alpha_C, \theta, D^*, D) &\propto \prod_{j=1}^{N} \phi\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right) \\ &= \exp\left\{\frac{1}{2}\sum_{j=1}^{N}\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right)^2\right\} \end{aligned}$$

Therefore,

$$\gamma/. \sim N(\left(Z'Z\right)^{-1}(Z'(D^* - \theta\alpha_C)), \left(Z'Z\right)^{-1}) \tag{E.4}$$

$$\begin{aligned} f(\alpha_C/\gamma, D^*, D) &\propto \exp\left\{-\frac{1}{10}(\alpha_C)^2\right\} \prod_{j=1}^{N} \phi\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right) \\ &= \exp\left\{\frac{1}{2}\sum_{j=1}^{N}\left(D_j^* - Z_j\gamma - \alpha_C\theta_j\right)^2 - \frac{1}{2}\frac{1}{10}(\alpha_C)^2\right\} \end{aligned}$$

Following the arguments presented in (E.2):

$$\alpha_C/. \sim N(\widehat{\alpha}_C, \overline{\Sigma}_C) \tag{E.5}$$

where $\overline{\Sigma}_C = \left(\theta^2 + \frac{1}{10}\right)^{-1}$ and $\widehat{\alpha}_C = \overline{\Sigma}_C(\theta\left(D^* - Z\gamma\right))$.

Lastly,

$$f(D^*/\alpha_C,\theta,\gamma,D) \propto \left[1(D_j^*>0)\phi\left(D_j^*-Z_j\gamma-\alpha_C\theta_j\right)\right]^{D_j}\left[1(D_j^*<0)\phi\left(D_j^*-Z_j\gamma-\alpha_C\theta_j\right)\right]^{1-D_j}$$

Therefore, D^* can be sampled from:

$$D_j^* = \begin{cases} TN_{[0,\infty)}\left(Z_j\gamma+\alpha_v\theta_j,1\right) \text{ if } D_j=1\equiv S \\ TN_{(-\infty,0]}\left(Z_j\gamma+\alpha_v\theta_j,1\right) \text{ if } D_j=0\equiv N \end{cases} \tag{E.6}$$

4. Precisions:

$$\tau_i/. \sim G\left(\frac{\sum_{j:D=i}1}{2}+2,\left(\frac{1}{2}\sum_{j:D=i}(y_j-x_j\beta_i-\alpha_i\theta_j)^2\right)+1\right) \quad \text{for} \quad i=N,S \tag{E.7}$$

$$\tau_T/. \sim G\left(\frac{N}{2}+2,\left(\frac{1}{2}\sum_{j=1}^{N}(T_j-W_j\eta-\theta_j)^2\right)+1\right) \tag{E.8}$$

$$\tau_\theta/. \sim G\left(\frac{N}{2}+2,\sum_{j=1}^{N}\theta_j^2+1\right) \tag{E.9}$$

Lastly, the posterior for θ becomes:

$$\begin{aligned}
\theta_j/.,D_j=1 \quad &\propto \quad \tau_i^{1/2}\exp\left(-\tau_i\frac{1}{2}(Y_j-X_j\beta_i-\alpha_i\theta_j)^2\right) \\
&\quad\quad \tau_T^{1/2}\exp\left(-\tau_{u_{T1}}\frac{1}{2}(T_j-W_j\eta-\theta_j)^2\right)\exp\left(\frac{1}{2}\left(D_j^*-Z_j\gamma-\alpha_C\theta_j\right)^2\right) \\
&\quad\quad \tau_f^{1/2}\exp(-\tau_f\theta_j^2) \\
&\propto \quad \exp\left(\tau_i(Y_j-X_j\beta_i-\alpha_i\theta_j)^2\right)\exp\left(\tau_T(T_j-W_j\eta-\theta_j)^2\right) \\
&\quad\quad \exp\left(\left(D_j^*-Z_j\gamma-\alpha_C\theta_j\right)^2\right)\exp(\tau_f\theta_j^2) \\
&= \quad \exp\left(\tau_i(\theta_j^*-\alpha_i\theta_j)^2+\tau_{u_{T1}}(\theta_j^{**}-\theta_j)^2+\left(\theta_j^{***}-\alpha_C\theta_j\right)^2+\tau_\theta\theta_j^2\right),
\end{aligned}$$

where $\theta_j^*=Y_j-X_j\beta_i$, $\theta_j^{**}=T_j-W_j\eta, \theta_j^{***}=D_j^*-Z_j\gamma$. Notice that:

$$\begin{aligned}
\alpha_i^2\tau_i\left(\frac{1}{\alpha_i}\theta_j^*-\theta_j\right)^2+\alpha_t^2\tau_T(\theta_j^{**}-\theta_j)^2 \quad &= \quad (\alpha_i^2\tau_{u_i}+\alpha_t^2\tau_{u_{T1}})\left(\theta_j-\frac{\alpha_i\tau_{u_i}\theta_j^*+\alpha_t\tau_{u_{T1}}\theta_j^{**}}{(\alpha_i^2\tau_{u_i}+\alpha_t^2\tau_{u_{T1}})}\right)^2 \\
&\quad\quad +\frac{(\alpha_i^2\tau_{u_i})(\alpha_t^2\tau_{u_{T1}})}{(\alpha_i^2\tau_{u_i}+\alpha_t^2\tau_{u_{T1}})}\left(\frac{1}{\alpha_i}\theta_j^*-\frac{1}{\alpha_t}\theta_j^{**}\right)^2,
\end{aligned}$$

and

$$\alpha_C^2\left(\frac{1}{\alpha_C}\theta_j^{***}-\theta_j\right)^2+\tau_f\theta_j^2=(\alpha_C^2+\tau_f)\left(\theta_j-\frac{\alpha_C\theta_j^{***}}{(\alpha_C^2+\tau_f)}\right)^2-\frac{\alpha_C^2\tau_f}{(\alpha_C^2+\tau_f)}\left(-\frac{1}{\alpha_C}\theta_j^{***}\right)^2.$$

Thus,

$$\exp\left(\tau_i(\theta_j^* - \alpha_i\theta_j)^2 + \tau_T(\theta_j^{**} - \theta_j)^2 + \left(\theta_j^{***} - \alpha_C\theta_j\right)^2 + \tau_f\theta_j^2\right) \propto$$

$$\exp\left((\alpha_i^2\tau_i + \tau_T)\left(\theta_j - \frac{\alpha_i\tau_i\theta_j^* + \tau_T\theta_j^{**}}{(\alpha_i^2\tau_i + \tau_T)}\right)^2 + (\alpha_C^2 + \tau_f)\left(\theta_j - \frac{\alpha_C\theta_j^{***}}{(\alpha_C^2 + \tau_f)}\right).^2\right)$$

And since:

$$\exp\left(\tau_i(\theta_j^* - \alpha_i\theta_j)^2 + \tau_T(\theta_j^{**} - \theta_j)^2 + \left(\theta_j^{***} - \alpha_C\theta_j\right)^2 + \tau_f\theta_j^2\right) \propto$$

$$\exp\left(-\frac{1}{2}\left(\alpha_i^2\tau_i + \tau_T + \alpha C + \tau_f\right)\left(\theta_j - \frac{\alpha_i\tau_i\theta_j^* + \tau_T\theta_j^{**} + \alpha_C\theta_j^{***}}{(\alpha_i^2\tau_i + \tau_{u_{T1}} + \alpha_C^2 + \tau_f)}\right)^2\right),$$

yielding:

$$\theta_j/., D_j = i \sim N\left(\frac{\alpha_i\tau_{u_i}\theta_j^* + \alpha_t\tau_{u_{T1}}\theta_j^{**}\alpha_v\theta_j^{***}}{(\alpha_i^2\tau_{u_i} + \alpha_t^2\tau_{u_{T1}} + \alpha_v^2 + \tau_\theta)}, \left(\alpha_i^2\tau_{u_i} + \alpha_t^2\tau_{u_{T1}} + \alpha_v^2 + \tau_\theta\right)^{-1}\right). \tag{E.10}$$

The Gibbs sampling procedure becomes:

1. Choose initial values for the parameters, and an arbitrary first draw for the factor. For example, $\theta^{(m)} \sim N(0,1)$)

For $m = 1, M$

1. Sample $D_j^{*(m)}$ for $j = 1, ..., N$ according to (E.6)
2. Sample $\theta_j^{(m)}$ for $j = 1, ..., N$ according to (E.10)
3. Sample $\beta_i^{(m)}$ $(i = S, N)$ according to (E.1)
4. Sample $\alpha_i^{(m)}$ $(i = S, N)$ according to (E.2)
5. Sample $\eta^{(m)}$ according to (E.3)
6. Sample $\gamma^{(m)}$ according to (E.4)
7. Sample $\alpha_C^{(m)}$ according to (E.5)
8. Sample $\tau_i^{(m)}$ $(i = 1, 2)$ according to (E.7)
9. Sample $\tau_T^{(m)}$ according to (E.8)
10. Sample $\tau_f^{(m)}$ according to (E.9)

Iterate over m until convergence.

F Estimated Model Parameters

Table F.1: Measurement System Loadings: Women

	BY SE	F1 SE	BY Math	F1 Math	GPA	Control	Motivation	Action
Constant	-0.44	-0.52	-1.06	-1.24	-0.35	-0.25	0.16	-0.11
	(0.10)	(0.15)	(0.12)	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)
Both Parents	0.09	0.10	0.13	0.12	0.20	0.10	0.05	0.10
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
HH Income	-0.01	-0.00	0.10	0.13	-0.07	0.04	-0.05	-0.03
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Parental Education	0.01	0.02	0.06	0.06	0.03	0.01	-0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Minority	0.02	-0.03	-0.62	-0.56	-0.51	0.08	0.02	0.10
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
θ_{SE}	1.00	0.52	0.06	0.06	0.00	0.33	0.00	0.09
	(0.00)	(0.02)	(0.02)	(0.02)	(0.00)	(0.02)	(0.00)	(0.03)
θ_{NC}	0.00	0.00	-0.05	-0.05	0.27	1.00	1.15	1.41
	(0.00)	(0.00)	(0.04)	(0.04)	(0.04)	(0.00)	(0.05)	(0.06)
θ_{M}	0.00	0.00	1.00	1.19	0.65	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)
Sample Size				2,615				

Note: Table F.1 presents the estimated parameters from the measurement system presented in Section 3.2. I obtain these estimates by simulating 500 values of parameters using the estimated posterior from the MCMC estimator. The 'Sample Size' row denotes the number of female students included in the ELS sample used to estimated the model. Standard errors in parenthesis.

Table F.2: Measurement System Loadings: Women

	BY SE	F1 SE	BY Math	F1 Math	GPA	Control	Motivation	Action
Constant	0.06	-0.32	-1.00	-0.75	-0.56	-0.37	-0.32	-0.39
	(0.10)	(0.17)	(0.14)	(0.13)	(0.16)	(0.16)	(0.18)	(0.17)
Both Parents	0.01	0.04	0.10	0.05	0.17	0.05	0.14	0.11
	(0.03)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)
HH Income	0.02	-0.05	0.09	0.15	-0.04	0.01	0.03	-0.12
	(0.03)	(0.06)	(0.05)	(0.04)	(0.06)	(0.05)	(0.06)	(0.06)
Parental Education	0.01	0.03	0.07	0.05	0.02	0.02	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Minority	-0.02	-0.10	-0.52	-0.52	-0.59	-0.01	0.05	-0.03
	(0.03)	(0.07)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)
θ_{SE}	1.00	0.54	-0.04	-0.10	0.00	0.52	0.00	0.07
	(0.00)	(0.02)	(0.03)	(0.03)	(0.00)	(0.02)	(0.00)	(0.04)
θ_{NC}	0.00	0.00	-0.22	-0.21	0.39	1.00	1.77	2.12
	(0.00)	(0.00)	(0.06)	(0.06)	(0.07)	(0.00)	(0.10)	(0.12)
θ_M	0.00	0.00	1.00	1.15	0.55	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)
Sample Size				1,984				

Note: Table F.2 presents the estimated parameters from the measurement system presented in Section 3.2. I obtain these estimates by simulating 500 values of parameters using the estimated posterior from the MCMC estimator. The 'Sample Size' row denotes the number of male students included in the ELS sample used to estimated the model. Standard errors in parenthesis.

Table F.3: Estimated Loadings in Major Choice Equations

	Male			Female			P-Value		
	θ_M (1)	θ_{SE} (2)	θ_{NC} (3)	θ_M (4)	θ_{SE} (5)	θ_{NC} (6)	θ_M (7)	θ_{SE} (8)	θ_{NC} (9)
Panel A. Initial Major Choice									
STEM	0.443	0.344	-0.365	0.482	0.305	-0.121	0.763	0.740	0.229
	(0.078)	(0.085)	(0.151)	(0.103)	(0.081)	(0.135)			
Life Sciences	0.454	0.002	0.308	0.255	0.190	0.122	0.134	0.135	0.434
	(0.102)	(0.105)	(0.206)	(0.085)	(0.069)	(0.119)			
Business	-0.029	0.072	0.167	-0.096	0.298	-0.263	0.493	0.020	0.013
	(0.067)	(0.074)	(0.140)	(0.071)	(0.063)	(0.101)			
Health	0.058	-0.025	0.569	-0.126	0.133	-0.045	0.129	0.231	0.024
	(0.104)	(0.118)	(0.255)	(0.062)	(0.059)	(0.094)			
Panel B. College Continuation									
STEM	0.464	-0.174	0.472	1.126	0.248	0.444	0.082	0.204	0.957
	(0.143)	(0.154)	(0.264)	(0.353)	(0.294)	(0.446)			
Life Sciences	0.701	-0.299	0.309	0.836	-0.328	0.664	0.730	0.939	0.631
	(0.286)	(0.326)	(0.665)	(0.266)	(0.197)	(0.321)			
Business	0.266	0.027	0.149	0.909	-0.391	0.627	0.002	0.066	0.265
	(0.128)	(0.150)	(0.342)	(0.169)	(0.171)	(0.259)			
Health	0.991	-0.629	0.886	0.525	0.078	0.037	0.397	0.370	0.625
	(0.533)	(0.781)	(1.730)	(0.134)	(0.110)	(0.174)			
Other	0.315	-0.106	0.388	0.496	-0.012	0.199	0.038	0.322	0.315
	(0.064)	(0.075)	(0.161)	(0.059)	(0.058)	(0.097)			
Panel C. Initial Major Completion									
STEM	0.403	0.068	-0.020	0.534	0.437	-0.801	0.712	0.237	0.136
	(0.152)	(0.149)	(0.268)	(0.320)	(0.274)	(0.450)			
Life Sciences	0.959	-0.406	0.846	0.352	0.142	0.294	0.102	0.079	0.361
	(0.303)	(0.264)	(0.537)	(0.214)	(0.166)	(0.276)			
Business	0.285	-0.196	-0.095	0.122	0.180	-0.182	0.436	0.041	0.802
	(0.131)	(0.124)	(0.265)	(0.163)	(0.135)	(0.224)			
Health	0.782	-2.039	3.371	0.166	-0.226	0.311	0.556	0.240	0.234
	(1.038)	(1.537)	(2.563)	(0.138)	(0.112)	(0.167)			
Other	-0.102	-0.173	0.361	-0.038	-0.118	0.255	0.547	0.568	0.579
	(0.078)	(0.077)	(0.164)	(0.072)	(0.058)	(0.098)			

Note: Table F.3 displays the estimated factor loadings in college major choice (Panel A), college continuation decision (Panel B) and initial major completion (Panel C) for men (columns 1-3) and women (columns 4-6) across the latent math ability, math self-efficacy and non-cognitive ability factors. The last three columns present the p-values from a test of equality of the factor loadings for males ($\boldsymbol{\alpha^m}$) and females ($\boldsymbol{\alpha^f}$) where $H_0 : \boldsymbol{\alpha^m} = \boldsymbol{\alpha^f}$ and $H_1 : \boldsymbol{\alpha^m} \neq \boldsymbol{\alpha^f}$.

Table F.4: Estimated Loadings in Employment Equations

	Male			Female			P-Value		
	θ_M	θ_{SE}	θ_{NC}	θ_M	θ_{SE}	θ_{NC}	θ_M	θ_{SE}	θ_{NC}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Initial Major Completers									
STEM	0.035	0.358	-0.510	-0.514	-0.412	0.768	0.457	0.099	0.124
	(0.220)	(0.203)	(0.364)	(0.704)	(0.420)	(0.746)			
Life Sciences	-1.725	1.008	-2.256	-0.572	0.138	-0.061	0.173	0.169	0.091
	(0.766)	(0.586)	(1.231)	(0.357)	(0.236)	(0.415)			
Business	0.029	-0.171	0.658	0.064	0.162	-0.095	0.905	0.210	0.101
	(0.190)	(0.185)	(0.361)	(0.224)	(0.190)	(0.284)			
Health	-0.162	0.408	0.096	0.192	0.145	-0.189	0.906	0.937	0.925
	(3.000)	(3.342)	(3.005)	(0.225)	(0.181)	(0.293)			
Panel B. Major Switchers									
STEM	0.388	0.075	-0.127	0.725	6.758	-4.957	0.870	0.003	0.054
	(0.258)	(0.273)	(0.540)	(2.048)	(2.203)	(2.443)			
Life Sciences	0.102	-0.258	1.227	0.936	0.225	-3.422	0.406	0.582	0.098
	(0.700)	(0.596)	(0.928)	(0.720)	(0.642)	(2.646)			
Business	-0.151	-0.328	-0.350	0.878	-0.416	1.331	0.109	0.874	0.142
	(0.289)	(0.279)	(0.783)	(0.573)	(0.477)	(0.835)			
Health	1.671	-2.223	2.288	0.363	-0.460	0.120	0.148	0.131	0.302
	(0.872)	(1.149)	(2.081)	(0.239)	(0.208)	(0.277)			
Other	-0.167	0.306	-0.611	-0.109	0.037	-0.077	0.804	0.271	0.257
	(0.168)	(0.192)	(0.396)	(0.163)	(0.151)	(0.254)			
Panel C. College Dropouts									
STEM	0.082	1.193	-4.867	3.366	-1.597	0.219	0.178	0.171	0.145
	(0.493)	(0.853)	(2.139)	(2.387)	(1.848)	(2.752)			
Life Sciences	3.019	3.299	-1.624	5.059	-4.170	-2.151	0.550	0.037	0.906
	(2.544)	(2.868)	(3.567)	(2.269)	(2.126)	(2.697)			
Business	-0.309	-0.020	0.137	1.655	-1.485	1.298	0.088	0.286	0.626
	(0.536)	(0.786)	(2.021)	(1.017)	(1.125)	(1.266)			
Health	-0.419	3.270	2.176	0.665	0.096	0.066	0.685	0.337	0.505
	(2.661)	(3.291)	(3.139)	(0.278)	(0.255)	(0.387)			
Other	0.221	-0.122	-0.424	0.117	0.133	0.163	0.537	0.174	0.107
	(0.122)	(0.144)	(0.295)	(0.116)	(0.120)	(0.213)			

Note: Table F.4 displays the estimated factor loadings in the employment choice equations for men (columns 1-3) and women (columns 4-6) across the latent math ability, math self-efficacy and non-cognitive ability factors. The last three columns present the p-values from a test of equality of the factor loadings for males ($\boldsymbol{\alpha^m}$) and females ($\boldsymbol{\alpha^f}$) where $H_0 : \boldsymbol{\alpha^m} = \boldsymbol{\alpha^f}$ and $H_1 : \boldsymbol{\alpha^m} \neq \boldsymbol{\alpha^f}$. Panel A presents the loadings for initial major completers, Panel B for initial major switchers and Panel C for college dropouts.

Table F.5: Estimated Loadings in Wage Equations

	Male			Female			P-Value		
	θ_M	θ_{SE}	θ_{NC}	θ_M	θ_{SE}	θ_{NC}	θ_M	θ_{SE}	θ_{NC}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Initial Major Completers									
STEM	0.082	-0.048	-0.023	0.238	0.035	-0.172	0.512	0.678	0.610
	(0.065)	(0.065)	(0.102)	(0.229)	(0.189)	(0.274)			
Life Sciences	-0.003	-0.144	0.076	-0.077	0.114	0.044	0.785	0.226	0.951
	(0.216)	(0.180)	(0.482)	(0.164)	(0.114)	(0.203)			
Business	-0.047	-0.006	0.082	0.094	-0.010	0.067	0.081	0.957	0.915
	(0.054)	(0.053)	(0.118)	(0.060)	(0.051)	(0.075)			
Health	0.378	-0.099	0.132	0.091	0.074	-0.028	0.493	0.773	0.855
	(0.411)	(0.597)	(0.867)	(0.081)	(0.053)	(0.096)			
Panel B. Major Switchers									
STEM	-0.059	0.095	-0.078	0.148	-0.174	0.208	0.481	0.353	0.492
	(0.179)	(0.137)	(0.254)	(0.233)	(0.255)	(0.330)			
Life Sciences	-0.094	0.169	0.348	0.087	-0.042	-0.032	0.655	0.494	0.438
	(0.370)	(0.279)	(0.443)	(0.166)	(0.132)	(0.210)			
Business	-0.090	-0.071	0.215	0.088	0.005	0.045	0.375	0.589	0.550
	(0.086)	(0.085)	(0.204)	(0.181)	(0.112)	(0.198)			
Health	-0.127	0.192	0.497	0.120	0.007	0.185	0.454	0.606	0.726
	(0.318)	(0.351)	(0.884)	(0.086)	(0.075)	(0.097)			
Other	0.027	-0.074	0.146	0.241	-0.038	0.128	0.016	0.675	0.900
	(0.063)	(0.066)	(0.113)	(0.062)	(0.055)	(0.087)			
Panel C. College Dropouts									
STEM	-0.125	0.012	0.117	0.000	0.000	0.000	0.344	0.933	0.632
	(0.132)	(0.143)	(0.244)	(0.000)	(0.000)	(0.000)			
Life Sciences	0.000	0.000	0.000	0.358	0.036	0.595	0.380	0.894	0.241
	(0.000)	(0.000)	(0.000)	(0.408)	(0.270)	(0.507)			
Business	-0.100	0.156	0.066	-0.040	-0.005	0.220	0.800	0.532	0.789
	(0.152)	(0.196)	(0.518)	(0.181)	(0.167)	(0.250)			
Health	0.406	0.133	0.489	0.284	-0.099	0.128	0.919	0.930	0.801
	(1.191)	(2.655)	(1.423)	(0.142)	(0.113)	(0.188)			
Other	-0.111	-0.026	-0.032	0.025	0.017	-0.027	0.062	0.610	0.975
	(0.053)	(0.070)	(0.137)	(0.050)	(0.047)	(0.080)			

Note: Table F.5 displays the estimated factor loadings in the hourly wages equations for men (columns 1-3) and women (columns 4-6) across the latent math ability, math self-efficacy and non-cognitive ability factors. The last three columns present the p-values from a test of equality of the factor loadings for males ($\boldsymbol{\alpha^m}$) and females ($\boldsymbol{\alpha^f}$) where $H_0 : \boldsymbol{\alpha^m} = \boldsymbol{\alpha^f}$ and $H_1 : \boldsymbol{\alpha^m} \neq \boldsymbol{\alpha^f}$. Panel A presents the loadings for initial major completers, Panel B for initial major switchers and Panel C for college dropouts.