Multidimensional Skills and Gender Differences in STEM Majors

Fernando Saltiel*

McGill University

This Draft: January 9, 2021
First Draft: October, 2018

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 426, 855 Sherbrooke St. West, Montreal, QC H3A 2T7. Sergio Urzua 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 Katharine Abraham, Joe Altonji, Peter Arcidiacono, Erich Battistin, Jim Berry, Peter Q. Blair, Raicho Bojilov, Michela Carlana, Andrew Dustan, Nicolas Figueroa, Camila Galindo, Rob Garlick, Joshua Goodman, Dan Hamermesh, Joe Hotz, David Jaume, John Kennan, Jeanne Lafortune, Fabian Lange, Joan Llull, Adrienne Lucas, Rosario Macera, Arnaud Maurel, Daniele Paserman, Tomas Rau, Jean-Marc Robin, Jorge Rodriguez, Cristian Sanchez, Seth Sanders, Miguel Sarzosa, Perihan Saygin, Modibo Sidibe, Isaac Sorkin, Jora Stixrud, Ashley Swanson, Lesley Turner, Ed Vytlacil and Matthew White for their useful comments and suggestions. I am indebted to seminar participants at the University of Maryland, World Bank's Africa Region Gender Innovation Lab, University of Delaware, Universidad de los Andes, Universidad Catolica, Universitat Autonoma Barcelona, Banco de Mexico, Duke University, McGill University, PUC-Rio, RIDGE Labor, Society for Economic Dynamics, Society for Labor Economics and the Southern Economics Association for their helpful feedback. 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, previous papers have not considered the importance of non-cognitive skills in driving students' college major choices, possibly resulting in an overestimation of the importance of preferences in shaping major choices.

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); Humphries et al. (2017); Heckman et al. (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, which captures students' perceived ability to perform math-related tasks. 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. I allow for these components to be correlated, relaxing the factor independence assumption imposed in previous work and fitting in with the recent literature on latent factors (Prada and Urzúa, 2017). Furthermore, as I estimate the model separately by gender, I can recover gender-specific correlations across latent skills. As such, I find that while math ability is positively correlated with self-efficacy, the correlation is far higher for men (0.503) than for women (0.310). This result thus indicates a relative 'lack' of high-skilled women who are confident in their math abilities vis-à-vis their male counterparts.

I find that 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 6.3% of those who are in the bottom self-efficacy decile enroll in STEM, whereas 14% 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 an increase in self-efficacy among the highest math-ability males substantially increases their STEM participation, as 37.5% 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 (σ) would increase their STEM enrollment rates by close to 50%, 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. 48% of male enrollees in the bottom self-efficacy decile complete a STEM degree, rising to 68% for those in the top quintile. On the other hand, while only 27% of female STEM enrollees in the bottom self-efficacy decile successfully complete a degree, the completion rate for those in the top quintile is almost three times as large, exceeding 67%. 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 ability (Heckman et al., 2006). All in all, a shortfall in math self-efficacy negatively affects 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 generalized Roy 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. Moreover, 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.[1] 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). I further find that 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 dropping out from college. I similarly find that all STEM enrollees would benefit from completing their initial majors rather than switching to a different degree. I remark that the returns to conditional STEM completion relative to both dropout and switching to a different major 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), indicates that policies focused on boosting self-efficacy could have a sizable impact on women's STEM participation rates. Using the estimated model parameters, I examine the impact of a policy increasing high-math-ability women's self-efficacy by 0.25 standard deviations. This intervention would increase women's STEM enrollment rates by 10 percent relative to baseline participation rates, with larger impacts for women at the top of the math ability distribution. This policy would also succeed in boosting graduation rates from math-intensive majors by 15-20 percent, as well. While boosting STEM participation rates may be worthwhile for non-pecuniary reasons, policymakers may also be interested in the labor market benefits arising from policy interventions. I analyze the effect of the self-efficacy intervention on women's hourly wages and small, yet positive impacts, in the range of 1%. Well-targeted skill development policies may thus help in reducing gender differences in STEM participation and in narrowing gaps in early-career labor market outcomes.

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; Beffy et al., 2012; Stinebrickner and Stinebrickner, 2014; Kinsler and Pavan, 2015; Arcidiacono et al., 2016; Humphries et al., 2017, 2019; Mourifie et al., 2020). I contribute to this literature by analyzing how multiple dimensions of students' non-cognitive skills affect their major choices and conditional completion

[1] 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).

outcomes, while documenting heterogeneous impacts by gender. Importantly, I present novel evidence showing that while mathematical self-efficacy affects the likelihood of STEM enrollment for both men and women, it strongly shapes STEM completion only for female students. As such, this paper further fits in with an extensive literature examining the drivers of gender differences in major choices. These papers have analyzed various determinants of the factors contributing to these differences, including the importance of high school preparation (Joensen and Nielsen, 2016; Shi, 2018; Aucejo et al., 2019; Card and Payne, 2021), test scores and cognitive skills (Turner and Bowen, 1999; Dickson, 2010; Speer, 2017; Jiang, 2018; Astorne-Figari and Speer, 2019), preferences and beliefs (Zafar, 2013; Wiswall and Zafar, 2015; Kugler et al., 2017; Wiswall and Zafar, 2018; Ahn et al., 2019; Delaney and Devereux, 2019; Bordón et al., 2020), 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 of a particular dimension of students' non-cognitive skills — math self-efficacy — in shaping differential STEM enrollment by gender and by showing its importance for STEM completion rates among women.

This paper further fits in with 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 (Beffy et al., 2012; Arcidiacono, 2004; Kinsler and Pavan, 2015; Humphries et al., 2017, 2019) and regression discontinuity designs (Hastings et al., 2013; Kirkeboen et al., 2016). I contribute to this literature by recovering the gender-specific returns to enrolling in STEM majors, importantly showing the returns vary significantly depending on the alternative major they are compared against. Moreover, as the structure of the model allows me to identify students' next-best choices, I estimate gender-specific returns to STEM majors vis-à-vis students' second-most preferred majors. I further document the returns to initial degree completion and show heterogeneous returns to majors across students' latent abilities, which further differ by gender.

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 reduced form evidence on the drivers and the returns to college majors. In Section 3, I introduce the generalized Roy model of college major choices, along with the measurement system and the estimation procedure. In Section 4, I present evidence on the correlation of the latent factors and on sorting patterns into initial majors and final educational outcomes for both men and women. In Section 5, I present evidence on the gender-specific returns to majors and on the conditional returns to completing such majors. In Section 6, I present the estimated impacts from a policy intervention aimed at boosting women's math self-efficacy. Lastly, 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. 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, labor market outcomes 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).[2] At the same time, ELS data also includes two measures of students' mathematical self-efficacy, measured in the baseline and first follow-up survey. Self-efficacy is defined as an "individual's judgment about being able to perform a particular activity" (Murphy and Alexander, 2000). The two measures are constructed directly from five questions measured on a four-point Likert scale using principal component analysis. The questions ask students to rate themselves on whether they think they can do an excellent job on math tests, understand difficult math texts, understand difficult math classes, do an excellent job on math assignments and whether the student can master math class skills.[3]

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

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

[3] To examine whether these variables capture different underlying constructs of the non-cognitive skill vector, I perform an exploratory factor analysis (EFA) using the five available measures in the ELS. I present the results in Figure A.1. Under the assumption of orthogonal factors, EFA indicates the existence of two factors, the first of which largely loads on the three 'traditional' non-cognitive skill measures, while the second encompasses both math self-efficacy measures as well as action control and control expectation. Meanwhile, instrumental motivation only loads on the first factor. As such, these results indicate the existence of two underlying non-cognitive skill factors, informing the structure of the measurement system introduced in Section 3.2.

(age 20). Nonetheless, the final sample includes students who do not graduate with a four-year degree, bachelor's recipients as well as students who have enrolled in or completed a graduate degree. 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.[4] Kahn and Ginther (2017) have shown that the STEM gender gap is largely driven by differences in math-intensive fields. I thus group majors into five categories, which include math-intensive STEM, life sciences, business, health, and the remaining majors.[5]

Lastly, 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. In particular, 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 take advantage of these variables to construct a measure of hourly wages for each individual in the sample.[6]

Sample Restrictions. 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.[7] 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.[8]

[4]Kinsler and Pavan (2015) group majors into business, science and others, Jiang (2018) follows a binary STEM classification, and Ransom (2020) aggregates them into STEM, business, social sciences, education and others.

[5]Math-intensive STEM fields include degrees in engineering, computer science, mathematics, statistics and physics. Life science majors encompasses biology and biology-related degrees. Business degrees includes majors in business, management and marketing. Health includes majors in clinical sciences and for health professionals. The "Other" group includes majors in social sciences, architecture, agriculture, arts, communication, journalism, education, English and foreign languages, family sciences, pre-Law, philosophy, religion, theology, psychology, public administration, liberal arts degrees, and individuals who had not yet declared an initial major.

[6]To avoid including part-time workers in the estimation of the returns to majors, I examine wage outcomes for individuals who reported having worked at least 500 hours in 2011. Results are robust to alternative hours worked restrictions. Nonetheless, the model outlined in Section 3 incorporates individuals' employment decisions.

[7]I observe the full set of observed skill measures for 60.4% of the sample. In Table A.2, I examine the relationship between students' observed characteristics and the likelihood of having a missing test score. These covariates are only weakly predictive of measure-non-response and insignificant across a number of measures. As such, 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 generalized Roy model is identified despite small differences in non-respondents' observed characteristics.

[8]Table A.1 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	STEM	Life Sciences	Business	Health	Other
Panel A. Women						
	(1)	(2)	(3)	(4)	(5)	(6)
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***	16.02***
Family Income (Log)	10.86	11.09	10.95	10.71*	10.95	10.66
Underrepresented Minority	0.18	0.22	0.20	0.19	0.20	0.22
Skill Measures						
Baseline Math Exam	-0.11	0.28	0.15	-0.19***	-0.28***	0.08***
HS GPA	0.12	0.29	0.39	0.08*	0.05**	0.29**
Baseline Math Self-Efficacy	-0.16	0.29	0.16	-0.04***	-0.15***	-0.13***
Non-Cognitive Skills (PCA)	0.02	0.21	0.33	-0.04**	0.02	0.23**
Educational Outcomes						
College Dropout	0.23	0.21	0.23	0.26	0.26	0.18
Complete Initial Major	0.61	0.55	0.44*	0.59	0.42**	0.77***
Labor Market Outcomes						
Employed	0.77	0.77	0.65**	0.82	0.77	0.75
Hourly Wages	21.16	23.06	18.79	22.63	25.70	21.34*
Observations	2,615	119	176	297	370	1,653
		4.5%	6.7%	11.4%	14.2%	63.2%
Panel B. Men						
	Full Sample	STEM	Life Sciences	Business	Health	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Background Characteristics						
Both Parents	0.83	0.83	0.82	0.87	0.87	0.85
Parental Education	15.91	15.99	16.50**	15.91	15.16***	16.41
Family Income (Log)	11.00	11.12	10.64**	10.89	10.71*	11.17
Underrepresented Minority	0.16	0.18	0.17	0.16	0.14	0.14
Skill Measures						
Baseline Math Exam	0.15	0.47	0.45	0.06***	-0.20***	0.35***
HS GPA	-0.16	0.13	0.21	-0.20***	0.05	-0.14***
Baseline Math Self-Efficacy	0.23	0.60	0.42*	0.17***	0.19***	0.27***
Non-Cognitive Skills (PCA)	-0.02	0.08	0.18	-0.02	0.17	0.14**
Educational Outcomes						
College Dropout	0.23	0.17	0.25**	0.18	0.17	0.18***
Complete Initial Major	0.56	0.62	0.46***	0.64	0.20***	0.78**
Labor Market Outcomes						
Employed	0.77	0.81	0.59***	0.85	0.71*	0.70**
Hourly Wages	21.36	23.66	19.57**	23.15	27.52*	20.61***
Observations	1,984	369	127	335	69	1,084
		18.6%	6.4%	16.9%	3.5%	54.7%

Notes: Table 1 presents summary statistics for the main sample of female and male four-year college enrollees considered in the paper. 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, showing that the majority of the sample comes from two parent families and the average surveyed parent has completed close to 16 years of schooling. Skill measures, standardized in the full sample, show that men outpace women in math exam performance, whereas women earn higher grades in high school (Goldin et al., 2006; Pope and Sydnor, 2010). At the same time, despite no discernible gender gaps in non-cognitive skills, I find significant differences in math self-efficacy, as men's self-efficacy exceeds that of women by 0.39 σ. By age 26, one-fourth of the sample has failed to complete a four-year degree and average major switching rates are higher for men than for women. Lastly, there are small differences in hourly wages, fitting in with previous evidence showing a small gender gap in the early career (Fortin, 2008; Blau and Kahn, 2017).

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 4.5% of women start in these majors, compared to 18.6% of men. Both men and women enrolled in STEM tend to come from higher income 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. On the other hand, these differences are not as stark when considering students' high school grades and their non-cognitive skills. In fact, both men and women exhibit far 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. In Table A.3, I additionally present results from a multinomial logit of major choices at enrollment, which further shows that math test scores and self-efficacy predict STEM enrollment for both men and women, even when conditioning for other skill measures and background characteristics.[9]

On average, a larger proportion of women complete their initial major than men, 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. In Table A.5, I examine the factors driving STEM completion among students initially enrolled in these majors. For both males and females, higher math test scores and high school grades strongly predict the likelihood of STEM completion. On the other hand, while women with higher math self-efficacy are more likely to finished a STEM degree — a one σ increase in this skill dimension is associated with an increased likelihood of completion by 10.1 percentage points (p-value = 0.131), the corresponding coefficient for men is negative and not significant.

Labor Market Outcomes. Table 1 shows that health fields are the highest-paid in the early career for both genders, fitting in with previous evidence presented in 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 'Other' degrees.[10] However,

[9]Using data from the UK, Aucejo et al. (2019) have previously found that women's relative advantage in verbal skills contributes to the gender gap in STEM. To examine the importance of not including reading skills in the model presented in Section 3, I re-estimate the multinomial logit of major choices including baseline reading test scores as a control variable (Table A.4). The results indicate that reading test scores do not predict major choices for women in the United States, yet lower the likelihood of STEM enrollment for men. Nonetheless, including reading skills in the analysis does not significantly change the importance of other skill measures as drivers of major choices.

[10]Male STEM enrollees earn higher wages than their peers in the life sciences, and the differences are significant.

as students sort into STEM based on their observed characteristics and skills, these wage differences do not capture the returns to such majors. I thus estimate an OLS regression to explore the returns to majors upon controlling for background characteristics and baseline skills. The second column shows larger relative returns to business and health majors vis-à-vis STEM degrees for women, whereas the male wage premia in STEM majors remains large and significant upon controlling for baseline skill measures.[11]

While these results represent a first approximation towards understanding the gender-specific returns to college majors, OLS regressions rely on a strong selection-on-observables assumption to recover the causal returns to major choices. This assumption is particularly strong in light of extensive evidence showing that observed test scores and skill measures capture latent abilities with significant error (Heckman et al., 2006, 2018). Moreover, OLS regressions cannot recover heterogeneous returns to major choices across the latent ability distribution (Altonji et al., 2016). Lastly, these results fail to account for potential selection into employment, an important consideration in the early career (Hamermesh and Donald, 2008). To address these concerns, I next introduce a discrete choice model which accounts for endogenous sorting into college major choices, final educational attainment and labor market outcomes for both 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 affect educational decisions and associated labor market outcomes. The skill measures available in ELS data allow me to recover three dimensions of individuals' latent abilities, encompassing their math ability, non-cognitive skills and their math self-efficacy through a measurement system of observed test scores. As such, the model follows a generalized Roy (1951) framework, fitting in with previous work by Heckman et al. (2006, 2018); Rodríguez et al. (2018); Humphries et al. (2019), allowing for individuals' choices and outcomes to depend both on their observed and unobserved characteristics. The model thus allows for essential heterogeneity in the returns to major choices and subsequent attainment by accounting for selection into schooling choices based on students' potential gains on their latent abilities.

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. This framework combines

[11] In the third and sixth columns, I additionally control for baseline reading scores, yet the estimated wage premia are not different from the estimated returns without including these test scores. The robustness of the estimated returns to the inclusion of reading skills fits in with recent evidence showing small, or even negative, returns to reading skill in the labor market (Arcidiacono, 2004; Kinsler and Pavan, 2015; Sanders, 2015).

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).[12] 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 multimdensional skills in shaping major choices and labor market outcomes, without invoking strong functional form assumptions. Importantly, I estimate the model separately for each gender to allow for differential sorting patterns by gender and to capture gender-specific 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 outlined in Section 2. Their major choice depends on both their observed characteristics and their latent ability ($\boldsymbol{\theta}$). Let $V_{i,m}^{G}$ be the utility for student i of gender G of starting in major m.[13] $V_{i,m}$ 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 thus 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 ε_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:

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

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), I similarly define the second-best major by: $N_{i,j} = \operatorname{argmax}_{m \in \mathcal{M}|m^*} \{V_{i,m}\}$ 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

[12] As discussed below, the model is identified through functional form assumptions and a matching-on-unobservables assumption (Heckman et al., 2016).

[13] The supra-index G refers to an person's gender, which can either be male m or female f. For notational simplicity, I omit the gender superscript in the rest of the Section.

to complete a degree, the model further incorporates the college completion margin. For a student 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, yet their peers who instead remain enrolled, they lastly 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 similarly 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) and the error term is similarly independent of observed and unobserved characteristics. In this context, 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.[14]

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

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) 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.[15] The structure of the model implies that $\boldsymbol{\theta_i}$ drives the cross-correlations of major choices and labor market outcomes outcomes, implying that OLS will recover biased estimates of the returns to college majors. As a result, identifying the distribution of $\boldsymbol{\theta}$ is of paramount importance in this context.

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, which may be classical or non-classical in nature (Carneiro et al., 2003). 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.[16] By estimating the model separately by gender, I can account or the existence of differential measurement error in test scores for men and women (Cattan, 2013).

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). 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}. I remark that 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] As such, I specify the following linear model for the two self-efficacy measures:

$$SE_{i,j} = \beta_j^{SE} \boldsymbol{X}_i^M + \alpha_j^{SE} \theta_{i,SE} + e_{i,j}^{SE} \tag{7}$$

where $\boldsymbol{SE}_i$ is the vector of observed math self-efficacy measures, $\boldsymbol{X}_i^M$ is a vector of exogenous

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

[16] As an an extensive literature has shown the importance of family, cultural and social factors in determining the evolution of ability through childhood, I consider the components of θ to be fixed by the time of college enrollment, but not fixed from birth.

[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.

control variables and $\boldsymbol{e}_i^{SE}$ represents the error term, which is independent across observed measures j, observed characteristics and θ_{SE}. I similarly posit a linear model for non-cognitive measures:

$$NC_{i,k} = \beta_k^{NC}\boldsymbol{X}_i^M + \alpha_k^{SE}\theta_{i,SE} + \gamma_k^{NC}\theta_{i,NC} + e_{i,k}^{NC} \tag{8}$$

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

$$M_{i,l} = \beta_l^M\boldsymbol{X}_i^M + \alpha_l^{SE}\theta_{i,SE} + \gamma_l^{NC}\theta_{i,NC} + \eta_l^M\theta_{i,M} + e_{i,l}^M \tag{9}$$

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 measurement system, from observed characteristics (X_i^M) and latent abilities ($\boldsymbol{\theta}$). Moreover, error terms in the measurement system are also independent from decisions and outcomes.[18]

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. In Appendix B, I show how the measurement system secures the identification of the distribution of $\boldsymbol{\theta}$ following identification arguments introduced in Carneiro et al. (2003); Hansen et al. (2004); Williams (2020). I show that 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]

3.3 Identification and Estimation

Model Identification. Carneiro et al. (2003); Heckman and Navarro (2007); 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.[20] 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 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

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

[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]Despite small differences in non-respondents' observed characteristics presented in Table A.2, Piatek and Pinger (2016) show the parameters in the chice equations should be equivalent for individuals with and without missing measures. In Table A.4, I present evidence from a multinomial logit of major choices for women and men, which shows that having missing test scores does not modify the underlying parameters in the choice equation.

market outcomes are independent conditional on all observed characteristics and latent ability (a 'matching-on-unobservables assumption') — secures model identification.

Model Implementation. Table A.8 presents the variables used in the implementation of the model. In the measurement system, I include information on students' race, their family composition, parental education and family income as the set of exogenous control variables. Moreover, in equations (7) and (9), I additionally control for parents' responses to two questions measuring whether they believe that 'people can learn to be good at math' and whether 'people must be born with the ability to be good at math' to examine whether parents' math attitudes can influence students' math performance and self-efficacy. I include the full set of control variables in the educational choice equations (1)-(3), in the employment decision as well as in the wage equation.

I estimate the model separately by gender. Since latent factors have no location of their own, I assume that the vector of unobserved abilities for males ($\boldsymbol{\theta}^m$) and females ($\boldsymbol{\theta}^f$) is a random variable with mean zero ($E(\boldsymbol{\theta}^m) = E(\boldsymbol{\theta}^f) = 0$).[21] As noted above, 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 by using a mixture of two normal distributions with means $(\mu_{1,f}, \mu_{2,f})$, probabilities $(p_{1,f}, p_{2,f})$, with $p_{1,f} + p_{2,f} = 1$, and variances $((\sigma_{1,f})^2, (\sigma_{2,f})^2)$ as follows:

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

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 $m \in \mathcal{M}$ in $\mathbf{T}_i$. Let Ψ be the vector of model parameters. While the model is identified non-parametrically, I estimate the model using normal distributions for the idiosyncratic shocks in the measurement system, initial major choice, college continuation decision, final major choice, employment decision and in the wage equation.[22] Given the independence assumptions invoked above, the likelihood for a set of I individuals is given by:

$$\begin{aligned}\mathcal{L}(\Psi \mid \cdot) \;\; = \;\; & \prod_{i\in\mathcal{I}} \Bigg[\int_{\boldsymbol{\theta}} \prod_{m\in\mathcal{M}} f(T_{im} \mid X_i^M, \boldsymbol{\theta}) \prod_{s\in\mathcal{S}} \Big\{ P(D_{is} = 1 \mid \mathbf{X}_i, \boldsymbol{\theta}) \left[f(Y_{is} \mid X_{is}^Y, \boldsymbol{\theta}) P(E_{is} = 1 \mid X_{is}^E, \boldsymbol{\theta}) \right]^{E_{is}} \\ & \left[1 - P(E_{is} = 1 \mid X_{is}^E, \boldsymbol{\theta})\right]^{1-E_{is}} \Big\}^{D_{is}} dF_{\boldsymbol{\theta}}(.) \Bigg]\end{aligned}$$

[21] Urzua (2008); Attanasio et al. (2020) show how to identify differences in the means of latent abilities across different groups, yet their identification arguments only apply under a dedicated measurement system. As such, the triangular measurement system introduced above does not allow me to recover differences in the mean of $\boldsymbol{\theta}$. While this assumption implies that I cannot discern the extent to which gender differences in latent abilities contribute to the gender gap in major choices, the model allows me to present extensive evidence on gender-specific sorting-into-majors based on multidimensional skills and to recover gender-specific returns to majors. I lastly remark that despite this assumption, I can still capture distributional differences in latent abilities.

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

where $f(T_m|\cdot)$ is the conditional density function of test score m, $f(Y_s|\cdot)$ 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).[23] I generate 500 draws from the estimated posterior distribution of the model parameters and simulate 200 samples where each simulated sample draws from the posterior of the estimated model parameters, yielding 523,000 and 396,800 simulated observations for women and men, respectively. Inference follows standard Bayesian arguments, as the Bernstein-von Mises theorem allows me to obtain the associated standard errors exploiting the standard deviations computed from these draws.

Goodness of Fit. To examine the validity of the discrete choice model in matching observed educational choices, I conduct various goodness of fit tests. First, in Figure A.2, I contrast students' observed initial major choices by gender against those simulated in the model. The model accurately predicts major choices by gender, with the majority of students in 'Other' majors and men outpacing women in STEM majors. Figure A.3 further shows that the model model correctly matches the share of female and male students who complete their initial major. I conduct similar goodness of fit tests to examine how well the model matches labor market outcomes. In Figure A.4, I show that the model closely matches employment shares for both male and female students who completed their initial major — as observed and simulated employment rates are in the 75-80% range across majors. Lastly, Figure A.5 shows that the model closely predicts the average hourly wages of male and female initial major completers who were employed in 2011.

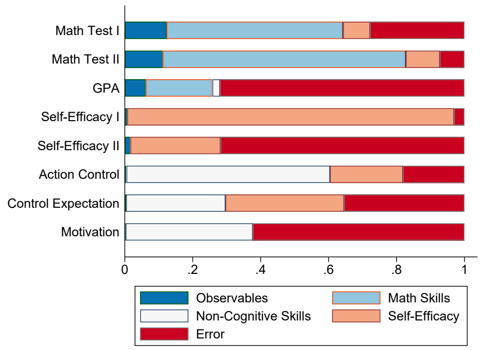
4 Latent Skills and College Major Choices

4.1 Measurement System and Latent Skills

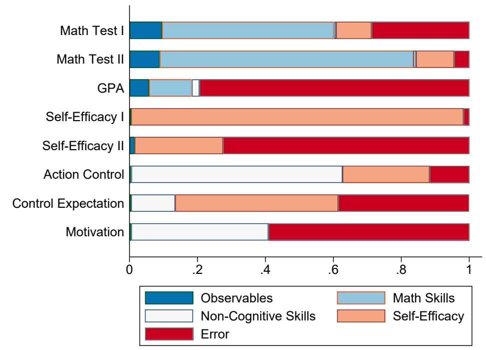
Tables D.1 and D.2 present the estimated coefficients from equations (7)-(9) for women and men, respectively. Women from two parent families are more likely to earn higher math test scores, have higher math self-efficacy and score higher on the three 'traditional' non-cognitive skill measures. Similar results are found for men, yet the coefficients are smaller in magnitude. Importantly, I do not find evidence that parental attitudes towards learning math significantly affect both female and male students' math performance or their self-efficacy. The loadings on the latent factors show that students with higher θ_{SE} get higher math test scores, whereas non-cognitive skills have an insignificant (or even negative, for males) impact on math performance. I lastly remark that 'traditional' non-cognitive skill measures load on students' latent self-efficacy, as well, remarking the importance of considering multiple dimensions of students' non-cognitive abilities.

[23]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, I make 500 draws from the posterior distribution of estimated model parameters to compute the mean and the standard errors of the parameters of interest. Appendix C describes the estimation algorithm in detail.

Figure 1: Variance Decomposition of Measurement System by Gender



(a) Variance Decomposition: Women



(b) Variance Decomposition: Men

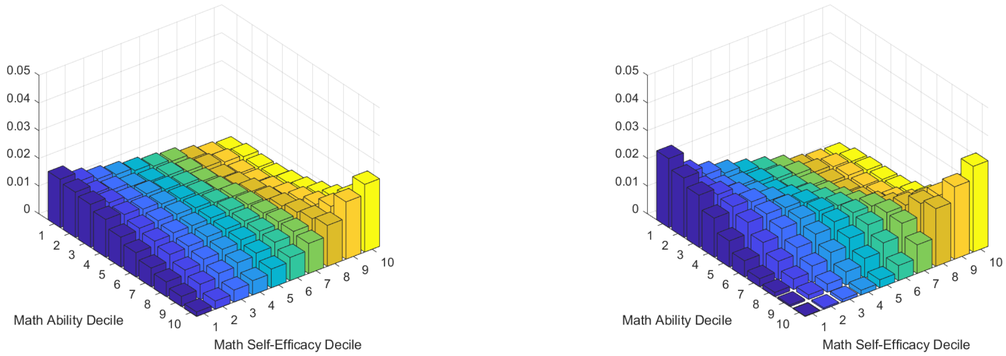
Note: Figure 1 presents the contribution of observed characteristics, latent abilities ($\boldsymbol{\theta}$) and the error term to the variance of the observed skill measures considered in the model. The row Observables indicates the share of the variance of the measurement variables explained by the observed variables: parental education, family composition, household income and students' underrepresented status. For math measures and self-efficacy, I additionally consider parents' math attitudes. Each Ability bar indicates the share of the variance explained by each component of the latent ability vector. Finally, the label Error term 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.

To understand the relative contribution of students' background characteristics and their latent ability vector for each test score, I present a variance decomposition of the measurement system in Figure 1. For both math test scores and high school grades, the share explained by observable characteristics reaches 10 percent for both men and women. Observed characteristics explain a much smaller share of the variance in 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 50 and 70 % of the variance in achievement. Moreover, 15% of the variance of high school GPA is explained by θ_M with an additional 5% explained by θ_{NC}, confirming Borghans et al. (2008)'s finding that GPA is a function of both cognitive and non-cognitive skills. Meanwhile, a sizable share of the variance in observed self-efficacy measures is explained by θ_{SE}, accounting 35-90% of the variance. Lastly, 20-55% of the variance in students' action control and control expectation is explained by θ_{NC}, yet an additional share is accounted for by their math self-efficacy, fitting in with the evidence presented in Figure A.1.[24] All in all, this evidence supports the argument that test scores cannot be equated with latent ability, as they are direct functions of background characteristics and capture distinct components of the ability vector.[25]

[24]Tables D.1-D.2 and Figure 1 show that the magnitude of the loadings and the variance decomposition of the measurement system is largely similar across genders, respectively. These results thus provide evidence of configural invariance (Cattan, 2013) — which requires for observed measures to be dedicated to the same unobserved ability component for both men and women.

[25]Figure A.6 presents the marginal densities of the latent factors for women and men. The estimated densities of $\boldsymbol{\theta}$ exhibit significant deviations from normality, remarking the importance of relaxing the assumption of normal

Figure 2: Correlation 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 2 presents the joint density of math ability and self-efficacy by gender, documenting the share of individuals pertaining to each joint decile of the two latent factor distributions. The first panel presents results for women and the second panel presents evidence for men.

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 0.503, far surpassing the 0.310 correlation for women. In Figure 2, I present the joint distribution of math ability and self-efficacy. A large share of men and women in their own-gender's top decile of θ_M are also in their the top math self-efficacy decile, yet while 30.9% of men in the top math ability decile are also in the top self-efficacy decile, the equivalent share is 24.3% for women. Furthermore, just 2.9% of men in the top math decile are below the median of the self-efficacy component, yet this is the case for 20.9% of women, confirming an over-representation of high-skilled women who lack confidence in their math ability.[26] Figure A.7 presents the joint distribution of the remaining latent factors. While the correlation between θ_M and θ_{NC} is lower for women (0.437) than for men (0.51), the difference is not as large as with θ_{SE}. I next consider the contribution of these differences in students' major choices.

4.2 Initial Major Choices

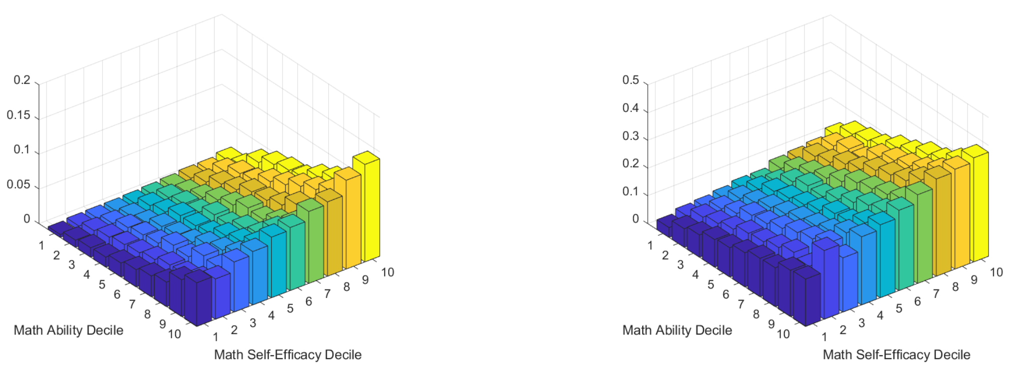
Using the estimated model parameters, I first analyze how students sort into initial majors based on their latent abilities. The first two panels of Figure A.8 show that both women and men sort into STEM majors based on their mathematical ability. In fact, women in STEM outpace their peers in life sciences by 0.299 σ in θ_M, and the difference reaches 0.66 σ compared to their counterparts

distributions for the latent factors.

[26] While examining the origins of the lack of high-skilled women who lack self-confidence in their math ability is beyond the scope of this paper, Carlana (2019) has recently found that teachers' gender stereotypes lower girls' subsequent performance and self-confidence in math with Lavy and Sand (2018) presenting similar evidence on the impacts of teachers on female students' test scores in Israel. These results thus suggest the lack of relative self-confidence for women may be a function of external influences.

in business. Similar patterns emerge for men, yet the extent of sorting on θ_M is more muted, as males in STEM just outpace their peers in life sciences and business by 0.069 and 0.582 σ in their math ability, respectively. Panels (c) and (d) further show that the θ_{SE} CDF of female and male students enrolled in STEM majors stochastically dominates the distribution of those in other majors, respectively. As such, these results indicate that students sort into STEM based both on their math ability as well as on their self-confidence in their math abilities. These patterns are not replicated for latent non-cognitive ability, as students in the life sciences have the highest θ_{NC} vis-à-vis their peers in other majors.[27]

Figure 3: STEM Enrollment Rates by θ_M and θ_{SE} by Gender



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

Note: Figure 3 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.

In light of the relative lack of high math ability women with high math self-efficacy, in Figure 3, I examine the relationship between students' latent skills and their STEM enrollment decisions. 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 (14 percent) than those in the middle joint decile (3.3 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 7.7 percentage points.[28] The second panel presents evidence for men, which shows that 34.7% 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 of the θ_{SE} to the top decile would similarly more than double their STEM participation, from 17.1% to 37.5%.[29] These results thus show that despite similar sorting-into-STEM patterns

[27]There are limited differences in male and female students' math ability, self-efficacy and non-cognitive skills across business, health and 'Other' major enrollees.

[28]These results remark the importance of considering additional skill dimensions when examining the determinants of major choices, as a one σ increase in θ_{NC} for women only increases the likelihood of STEM enrollment by 0.7 percentage points, whereas a corresponding increase in their self-efficacy would boost their STEM participation by 2.2 percentage points.

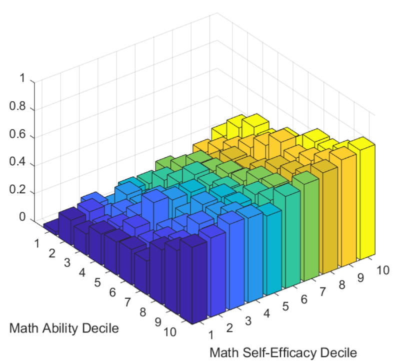
[29]A similar share of males and females in the top decile of the joint θ_M and θ_{SE} distribution enroll in the life

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

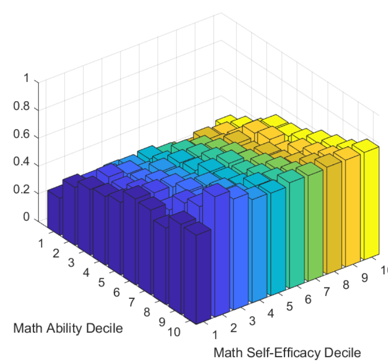
4.3 Final Major Choices

While 62% of male STEM enrollees end up succesfully completing a major in this field, this is the case for just 55% of their female counterparts. In Figure A.9, I examine how latent abilities shape final educational outcomes for initial STEM enrollees. For both women and men, higher math ability and self-efficacy substantially increase the likelihood of completing a STEM degree. Meanwhile, the last two panels show that while higher non-cognitive ability leads to an increased likelihood of college graduation (Heckman et al., 2006, 2018), it is not associated with a higher likelihood of completing a STEM degree.

Figure 4: STEM Completion Rates Among Initial Enrollees: by Gender



(a) STEM Completion Rates: Women



(b) STEM Completion Rates: Men

Note: Figure 4 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. The first panel presents results for women and the second panel presents evidence for men.

In Figure 4, 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 θ_M distribution increases the likelihood of STEM completion from 40.6% to 73.1%. Math self-efficacy is similarly important, as the corresponding increase from the middle to the top decile of the θ_{SE} distribution

sciences and in business (10-12%). The largest differences emerge in health fields (10.8% for women; 2.7% for men, see Altonji et al. (2016)) and in 'Other' majors, where female participation equals 50.9% compared to 37.3% for men.

would boost completion rates from 42.0% to 67.3%. The joint skill distribution of math ability and self-efficacy presents a similar story, as 79% of women in the top joint decile graduate with a STEM degree, yet this share drops to below 55% percent for those in the top problem solving decile and in the bottom of the θ_{SE} distribution.

The second panel of Figure 4 shows that these patterns are strikingly different for men, for whom self-efficacy plays a far smaller role in driving STEM degree completion. While 57.7% of STEM enrollees in the middle decile of the θ_{SE} distribution complete a degree after enrollment, this share rises only slightly 67.8% top self-efficacy decile. This result shows the importance of considering how different margins of ability differentially affect men and women's progress through majors in college. In particular, math self-efficacy play a critical role for women's exit from STEM, yet this margin has not received much attention in the existing literature. While the model presented in Section 3 does not directly account for the mechanisms through which low math self-efficacy drives STEM dropout, a female student with low self-efficacy may leave STEM upon getting a low grade in a STEM-based class. In fact, Ahn et al. (2019) have recently shown that since women place a higher value on grades than men, harsh grading policies in STEM classes contribute substantially to the gender gap in these majors.

Lastly, in Figure A.10, I analyze the importance of students' abilities in leading to degree completion among non-STEM enrollees. The results are strikingly different relative to the estimated importance of $\boldsymbol{\theta}$ for STEM enrollees. For instance, while θ_M increases the likelihood of staying in college, it does not have a differential impact on the likelihood of completing the initial degree. I similarly find that there is no discernible relationship between students' mathematical self-efficacy and initial-degree completion for non-STEM enrollees. This is the case for both females and males, as shown in panels (c) and (d), respectively. All in all, these findings show that latent non-cognitive skills and mathematical self-efficacy play different roles in leading to successful outcomes in higher education, further remarking the importance of considering multiple dimensions of skills in the analysis of major choices.

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. An extensive literature has estimated the returns to graduating from different majors. Altonji et al. (2012) highlight papers which have previously estimated gender-specific returns. A common empirical strategy, followed by Altonji (1993); Rumberger and Thomas (1993); Chevalier (2011); Webber (2014), among others, estimates a linear regression with controls for pre-college test scores. These papers find positive returns for women graduating from engineering, math and science degrees, relative to a degree in education. However, this empirical approach does not account for sorting into majors on unobserved characteristics, nor does it provide evidence on heterogeneous returns to majors across individuals' latent skills. Jiang (2018) advances the literature by estimating a Roy model of major choices, finding positive returns to STEM degrees. Her analysis contributes to the literature by differentiating test scores from latent abilities, yet the binary classification of STEM degrees does not incorporate heterogeneous returns across multiple major choices.[30] Humphries et al. (2017) address these issues using Swedish data, but they do not examine gender-specific returns. Importantly, these papers do not directly account for endogenous sorting-into-employment (Hamermesh and Donald, 2008), an important consideration in the early career in light of low full-time employment participation in the ELS sample.

In this context, 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.[31] Letting 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:

$$ATE^G_{S,m} \equiv \int \int E[Y^G_{i,S} - Y^G_{i,m} | \boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}] dF_{X,\boldsymbol{\theta}}(x, \underline{\boldsymbol{\theta}}) \;\; for \;\; m, f \in \{G\} \tag{10}$$

where $Y^G_{i,S} - Y^G_{i,m}$ captures the wage returns to starting in a STEM major relative to enrolling in 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 $\boldsymbol{\theta}$, depending on the returns to each component of skills in both leading to college graduation and in increasing labor market productivity. I thus consider the heterogeneous returns to STEM enrollment across the latent ability distribution in:

$$ATE^G_{S,m}(\underline{\theta}_M, \underline{\theta}_{SE}, \underline{\theta}_{NC}) \equiv E[Y^G_{i,S} - Y^G_{i,m} | \theta_C = \underline{\theta}_C, \theta_{SE} = \underline{\theta}_{SE}, \theta_{NC} = \underline{\theta}_{NC}] \tag{11}$$

Importantly, equation (11) allows me to understand the extent to which high-skilled women would benefit from STEM.

5.2 Estimated Returns to Majors

In the first panel of Table 2, I present the estimated returns to enrolling in a STEM major for women vis-à-vis the other four majors considered in the analysis. First, in Table 1, I had shown that hourly wages for women in STEM majors were 9.5% higher relative to those of their peers in the life sciences. The second column of Table 2 shows that the estimated ATE of STEM enrollment relative to the life sciences, estimated using simulated parameters from the model, equals 7.2%.

[30] These papers estimate the returns to completing a specific major, yet in a context of sequential major choices, the returns to major completion capture a different parameter than the returns to enrollment, as the latter incorporate the possibility that a student may not subsequently complete the major. In fact, when considering the benefits arising from a policy nudging students to enroll in a different major, policymakers should be interested in the latter parameter, which represents a linear combination of the wages of major completers, switchers and college dropouts.

[31] The estimated returns to majors presented in this paper are relevant only in the early career, thus not representing lifecycle returns to college majors. These parameters are informative for understanding whether raw wage gaps across majors are reflective of 'underlying' returns to such degrees or a function of sorting-into-majors.

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 not statistically significant, the returns relative to business majors are negative and significant, equaling -11.7%. The returns to enrolling in STEM relative to starting in a health-related fields 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. Across all alternative majors, I remark that the estimated ATE of STEM enrollment 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 across both their latent math ability and self-efficacy.

Table 2: Estimated Returns to STEM v. Alternative Majors by Gender

		Panel A. Women		
Estimate	Life Sciences	Business	Health	Other
ATE	0.072	-0.117	-0.250	-0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)
ATE (High θ_M)	0.129	-0.062	-0.266	0.039
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
ATE (Low θ_M)	0.003	-0.183	-0.230	-0.052
	(0.002)	(0.002)***	(0.002)***	(0.002)***
TT	0.085	-0.102	-0.269	-0.011
	(0.006)***	(0.005)***	(0.005)***	(0.005)**
TUT	0.073	-0.063	-0.230	-0.008
	(0.005)***	(0.003)***	(0.003)***	(0.001)***
		Panel B. Men		
Estimate	Life Sciences	Business	Health	Other
ATE	0.254	0.060	-0.167	0.217
	(0.001)***	(0.001)***	(0.003)***	(0.001)***
ATE (High θ_M)	0.308	0.143	-0.177	0.259
	(0.002)***	(0.002)***	(0.003)***	(0.002)***
ATE (Low θ_M)	0.190	-0.038	-0.155	0.168
	(0.002)***	(0.002)***	(0.005)***	(0.002)***
TT	0.292	0.103	-0.197	0.230
	(0.003)***	(0.003)***	(0.006)***	(0.003)***
TUT	0.242	0.047	-0.156	0.217
	(0.006)***	(0.003)***	(0.017)***	(0.002)***

Notes: Table 2 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 (10). ATE (High θ_M) and ATE (Low θ_M) present the estimated ATE to STEM majors for students above and below the gender-specific 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. Standard errors are in parenthesis: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

The first row of the second panel of Table 2 presents the corresponding returns to STEM enrollment for men. Enrolling in STEM delivers large 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 21% when compared to the life sciences or other majors. Why are the returns to enrolling in STEM significantly for women than they are for men? Differential STEM completion rates could explain part of the effect, as 55% of women finish this degree compared to 62% of men. However, fitting in with the evidence presented in Figure A.5, potential wages for enrolling in STEM majors for men are higher than they are for women. These results fit in with previous evidence presented in Altonji et al. (2012) who had shown that male math-intensive graduates earn higher wages than their female counterparts. Moreover, Eide (1994); Chevalier (2011) had similarly found higher returns to STEM majors for men than for women.[32]

I remark that the estimated returns may 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). For women, the TT and TUT parameters are presented in the last two rows of the first panel of Table 2, which are largely similar in magnitude to the estimated ATEs presented above. For men, I find that the estimated TT to STEM enrollment is larger than the corresponding ATE across all alternative majors, except with respect to health. In this context, Heckman et al. (2018) show that the difference in the estimated ATE and the observed wage difference across any two majors can be decomposed into the importance of selection bias — given by the difference in potential wages in major m for STEM enrollees vis-à-vis major m enrollees — and a 'sorting gains' parameter, which is positive if students who stand to have the largest gains from STEM enrollment are the ones enrolled in these majors. I present the results of the decomposition in Table A.10, finding that women are positively selected into STEM majors relative to all other available choices, whereas the extent of selection bias is smaller for men. On the other hand, since the sorting gains parameter varies in sign and is small in magnitude, I find limited evidence that students who stand to gain the most from STEM enrollment are those who pursued such majors.

While the returns to majors discussed so far are computed by integrating out the latent skill distribution, these may be heterogeneous across students' abilities. In Table 2, I further present the returns to STEM enrollment for students above and below the gender-specific θ_M median. I find that the returns to STEM for high-skilled women is generally larger than the estimated ATE. In Figure A.11, I further show that women in the top decile of the θ_M distribution earn positive returns to STEM enrollment when compared to otherwise starting in a life science, business or 'Other' major. The second panel of Table 2 similarly shows that higher-math-ability men would enjoy greater returns from STEM enrollment vis-à-vis their lower skilled peers.[33] These findings

[32] In Table A.9, I compare the model-based estimates of the returns to STEM enrollment to evidence from different reduced-form strategies, including different OLS approaches and nearest-neighbor matching. Across a number of majors, the estimated model-based ATE is significantly different to the estimated returns recovered under alternative strategies, remarking the importance of accounting for sorting-into-majors based on unobservable characteristics.

[33] Figure A.12 presents heterogeneous returns to STEM majors for men across the θ_M distribution.

thus fit in with an higher productivity of math ability for male and female students who pursue math-intensive degrees (Arcidiacono, 2004; Humphries et al., 2019).

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); Kirkeboen et al. (2016) and instead examine the returns to STEM majors for students who enrolled in such majors vis-à-vis their next-best option (defined in Section 3.1) as follows:

$$TT_{S,j} \equiv \int \int E[Y_{i,S} - Y_{i,j} | \boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}, D_{i,S} = 1, N_{i,j} = 1] dF_{X,\boldsymbol{\theta}}(x, \underline{\boldsymbol{\theta}}, D_{i,S} = 1, N_{i,j} = 1) \quad (12)$$

where $Y_{i,S} - Y_{i,j}$ recovers the returns to STEM enrollment compared to major j, and $TT_{S,j}$ denotes the returns to STEM majors for students enrolled in STEM ($D_{i,S} = 1$) whose second-best option is major j ($N_{i,j} = 1$).[34]

In Figure 5, 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 5.7%. 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 have benefited from remaining in STEM. For men, meanwhile, the estimated returns indicate that remaining enrolled in STEM instead of their next-best major would have been beneficial for almost all students, with increasing returns across the math ability distribution, as well.

Another policy-relevant margin in the context of unordered major choices 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.[35] The third panel of Figure 5 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 wage losses (-4.2%).[36] However, the $TT_{S,j}$ parameter becomes positive for female students in the top quintile of the θ_M distribution, fitting in with the positive returns to STEM enrollment for high math ability women shown above.

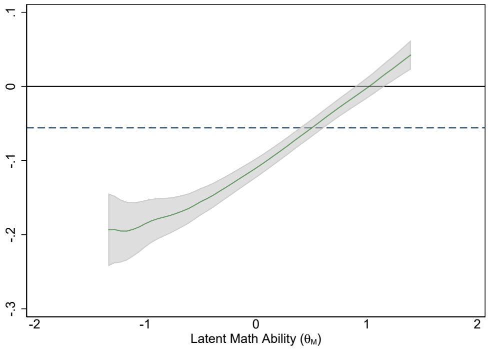
[34] For women who started in STEM, enrolling in a major in the 'Other' grouping would have been the preferred alternative for half of the sample (Table A.11). The remaining half would have instead chosen a major in the life sciences, health or business in equal shares. For men, I also find 55% of STEM enrollees had 'Other' majors as the next best option, and 24% of these students would have instead enrolled in STEM.

[35] This margin is captured through the following treatment-on-the-untreated parameter:

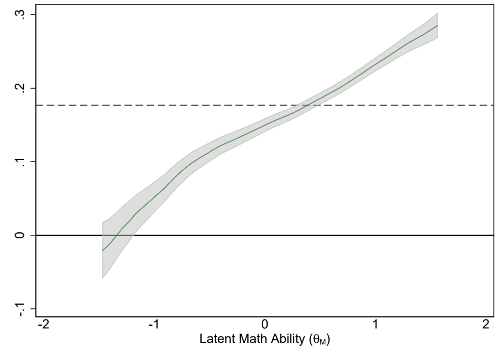
$$TUT_{S,j} \equiv \int \int E[Y_{i,S} - Y_{i,j} | \boldsymbol{X} = x, \boldsymbol{\theta} = \underline{\boldsymbol{\theta}}, D_{i,j} = 1, N_{i,S} = 1] dF_{X,\boldsymbol{\theta}}(x, \underline{\boldsymbol{\theta}}, D_{i,j} = 1, N_{i,S} = 1)$$

[36] Panel (d) presents the corresponding returns for men, indicating that males in other majors would have uniformly benefited from enrolling in STEM majors instead.

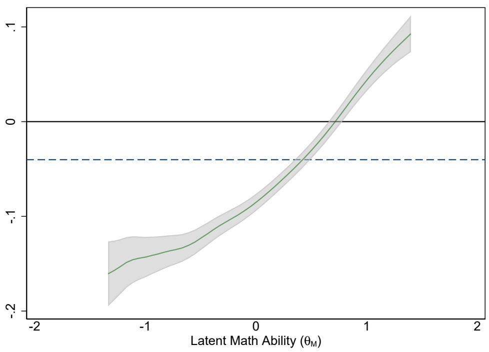
Figure 5: Heterogeneous Returns to STEM Enrollment Compared to Second-Best Option



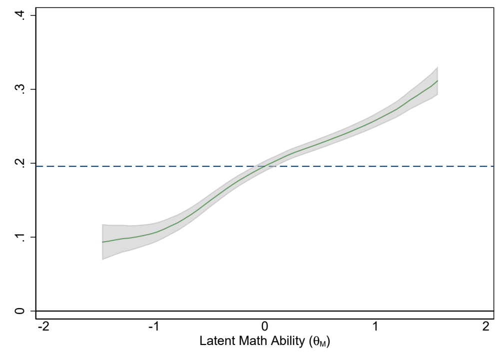
(a) TT Parameter: Women



(b) TT Parameter: Men



(c) TUT Parameter: Women



(d) TUT Parameter: Men

Note: The first panel of Figure 5 presents heterogeneous returns to STEM enrollment for women who actually enrolled in STEM relative to the next-best major (defined in equation (12)) across the θ_M distribution. The second panel presents corresponding returns for men. The third panel presents the estimated TUT parameter which, for women enrolled in other majors, recovers the returns to instead enrolling in STEM. The last panel presents the estimates of the same parameter for men.

5.3 Conditional Returns to Degree Completion

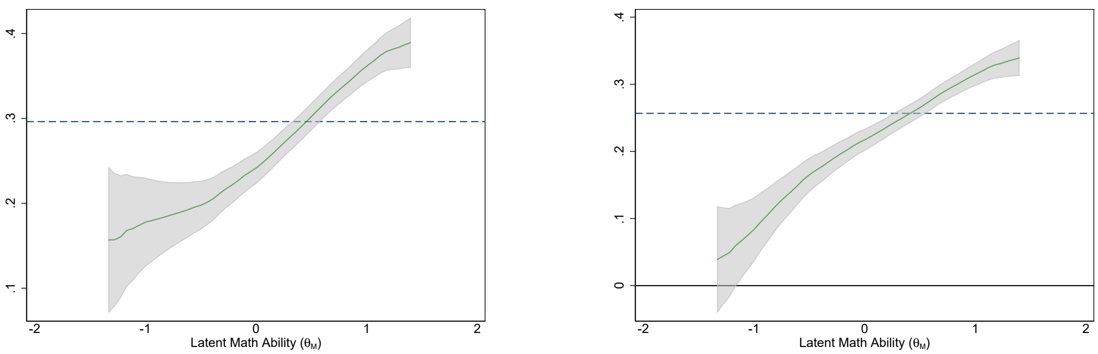
While the returns to STEM enrollment indicate that high-math-ability women would largely benefit from starting out in these majors, an open question remains as to whether remaining in these majors through completion would be necessarily beneficial. To this end, I define the conditional returns to completing initial major m relative to switching to a different major, or relative to dropping out of college altogether as follows:

$$\{ATE_{m,F,S}|D=m\} \equiv \int\int E[Y_{i,m,F} - Y_{i,m,S}|\boldsymbol{X}=x, \boldsymbol{\theta}=\underline{\boldsymbol{\theta}}, D=m]dF_{X,\boldsymbol{\theta}|D=m}(x,\underline{\boldsymbol{\theta}}) \tag{13}$$

$$\{ATE_{m,F,C_0}|D=m\} \equiv \int\int E[Y_{i,m,F} - Y_{i,m,C_0}|\boldsymbol{X}=x, \boldsymbol{\theta}=\underline{\boldsymbol{\theta}}, D=m]dF_{X,\boldsymbol{\theta}|D=m}(x,\underline{\boldsymbol{\theta}}) \tag{14}$$

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 6: Heterogeneous Returns to STEM Completion for Female STEM Enrollees



(a) Returns v. Switching

(b) Returns v. Dropout

Note: The first panel of Figure 6 presents heterogeneous returns to STEM completion relative to switching to a different major (equation (13)) for women. The second panel presents corresponding evidence relative to dropping out of college altogether (equation (14)).

In the first and second panel of Figure 6, 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 close to 30%. 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 similarly large (25.7%) and increasing in the math ability distribution. Figure A.13 similarly shows sizable returns for men from completing STEM degrees both relative to switching to a different major (35.6%) and to dropping out of college altogether (56.4%). Lastly, in Table A.12, I present the conditional returns to completion for students in other majors. First, all women would benefit from completing their initial (non-STEM) majors relative to dropping out. On the other hand, students in the life sciences, business, and 'Other' majors would earn positive returns from switching to other majors.[37]

All in all, these findings indicate that well-targeted policies aimed at increasing STEM completion among female enrollees would substantially increase their early-career wage outcomes. Given the positive returns to STEM majors for high math-skilled women presented in this section, I next examine whether different policy interventions could offer a pathway for increasing women's STEM participation along with delivering improved early-career labor market outcomes.

[37]The estimated returns for men show the returns to conditional completion vary both in sign and in magnitude relative to both switching and dropping out from college.

6 Policy Simulation: Math Self-Efficacy Increase

6.1 Simulated Intervention

Colleges across the country have implemented policies aimed at boosting students' STEM participation and subsequent completion rates, ranging from mentoring initiatives, STEM-program exposure, increased lab experience and summer preparation programs (Olson and Riordan, 2012). In this context, while I had shown that both math ability and self-efficacy are strong predictors of STEM enrollment and subsequent completion, non-cognitive skills are malleable through adolescence, which is not the case for cognitive abilities (Kautz et al., 2014). As such, policy interventions aimed at boosting women's math self-efficacy in high school could successfully impact their latent self-efficacy and thus increase STEM participation rates.[38] Previous papers in the psychology literature have found that different strategies can boost students' self-efficacy. For instance, Siegle and McCoach (2007) shows that a four-week course focused on improving high school math teachers' self-efficacy instructional strategies, which encompassed improving teacher feedback, establishing goals and presenting models of success, boosted students' math self-efficacy by 0.46 standard deviations. Cordero et al. (2010) and Betz and Schifano (2000) have similarly found positive effects of student-level self-efficacy interventions.[39] Nonetheless, since the existing psychology literature has not been precise about the feasibility of interventions of varying magnitudes, I examine how a policy which would boost women's self-efficacy by 0.25 standard deviations — which corresponds to less than the average gender gaps in baseline math self-efficacy presented in Table 1 — would affect their STEM participation rates.[40] In light of low enrollment rates among high-math-ability women, I focus the simulated intervention on female students above the θ_M median.

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.[41] 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:

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

[38] Huang (2013) shows that gender differences in mathematical self-efficacy grow substantially between middle-school and high-school, indicating that interventions aimed during this period may have a greater impact.

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

[40] Despite the positive correlation between θ_{SE} and θ_M, I assume that the simulated intervention would not jointly affect women's math ability. As a result, the estimated impacts presented below likely represent a lower bound on the potential effect arising from self-efficacy-based policies.

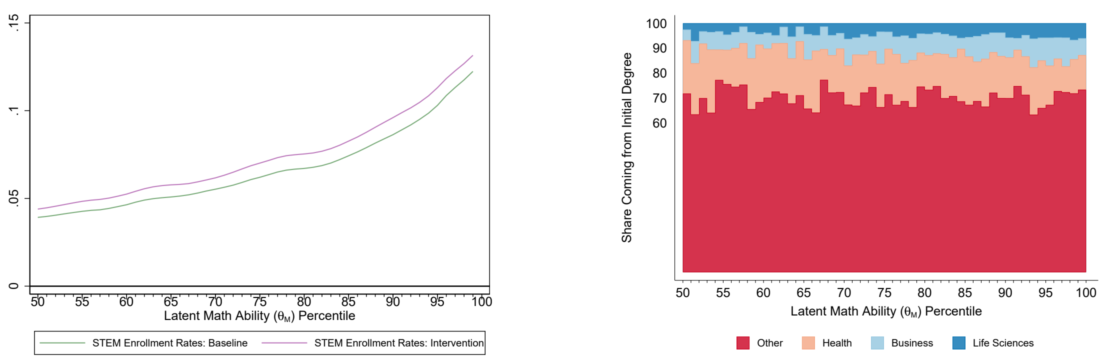
[41] The simulated impacts of the intervention should be understood through a partial equilibrium framework, as this exercise aims to illustrate the potential impacts of increasing women's self-efficacy on STEM participation and early-career labor market outcomes.

$$
\begin{aligned}
&E[Y_i(p') - Y_i | D_{i,s}(p') = 1, D_{i,s} = 1] \times \underbrace{P[D_{i,s}(p') = 1, D_{i,s} = 1]}_{STEM\ Enrollment\ Always-Takers} + \\
&E[Y_i(p') - Y_i | D_{i,s}(p') = 0, D_{i,s} = 0] \times \underbrace{P[D_{i,s}(p') = 0, D_{i,s} = 0]}_{STEM\ Enrollment\ Never-Takers} \qquad (15)
\end{aligned}
$$

where $D_{i,s}$ is a dummy variable which equals one for student i enrolled in STEM. Equation (15) 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, never-takers, and compliers, who are the students changing the enrollment decision due to the policy.[42] I lastly remark that the effect of these interventions may further vary across the $\boldsymbol{\theta}$ distribution.

6.2 Impacts on STEM Participation

Figure 7: Heterogeneous Returns to STEM Completion for Female STEM Enrollees



(a) Impacts on STEM Enrollment

(b) Initial Major Choices

Note: The first panel of Figure 7 presents the estimated impacts of the simulated self-efficacy intervention (a 0.25 σ boost in math self-efficacy for women above the θ_M median) 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 share of compliers across different percentiles of the θ_M distribution who would be switching out of different majors.

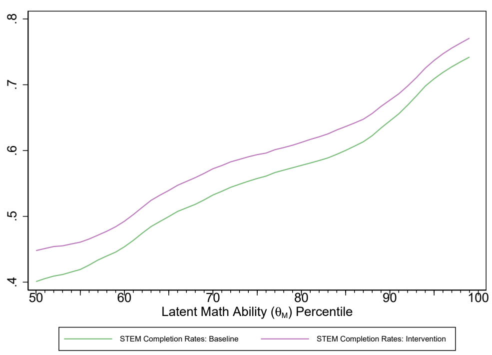
Impact on STEM Enrollment Rates. I first examine the effect of the self-efficacy intervention on women's STEM enrollment rates. The first panel of Figure 7 shows the share of high θ_M women who would enroll in STEM upon an increase of 0.25 σ in their self-efficacy. The simulated intervention would increase high θ_M women's STEM enrollment rates from 6.6% up to 7.3%, representing an 11% increase relative to baseline participation rates. The θ_{SE} increase would yield larger impacts for women in the top decile of the math ability distribution, as their participation rate in math-intensive majors would increase by a full percentage point, from 11.1% up to 12.1%.[43]

[42] I define response types by students' initial major decisions. Always-takers thus represent students who enroll in STEM both under baseline as well as in policy p'. Never-takers are those who do not enroll in STEM in either case. Compliers are those who choose to enroll in STEM as a function of p', yet had not done so in the baseline. The sorting-into-STEM patterns based on female students' latent self-efficacy implies there are no defiers in this context.

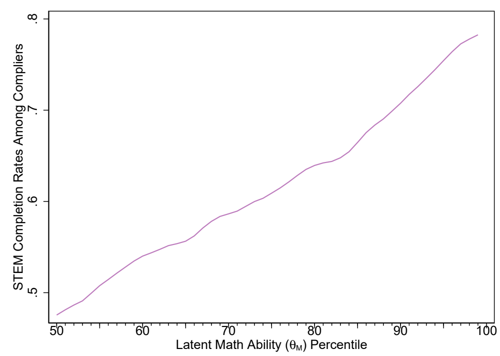
[43] In this setting, I further remark that depending on how preferences are formed during childhood, an early-life

The second panel shows the majors that 'compliers' would have chosen in absence of the policy, an important consideration for evaluating the labor market impacts of the simulated intervention in light of the returns presented in Section 5. I find that 60-70% of compliers would be drawn from 'Other' majors, with a largely constant share across the math ability distribution. An additional 15-20% of compliers would be switching from health majors, and an additional 5% would switch from business majors along with a negligible share coming from the life sciences. As such, the simulated self-efficacy intervention would be largely drawing students away from majors in the humanities, social sciences and education into math-intensive STEM fields.

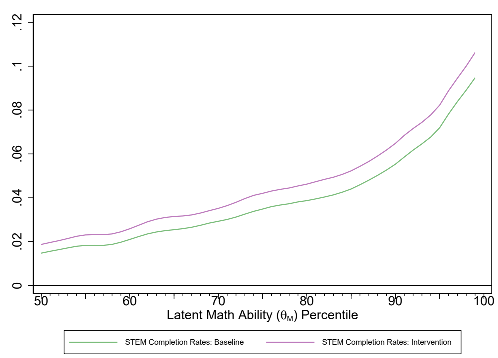
Figure 8: Estimted Impacts of Policy Intervention on STEM Completion Rates



(a) Impact on STEM Completion: Always-Takers



(b) Impact on STEM Completion: Compliers



(c) Impact on STEM Completion: Overall

Note: The first panel of Figure 8 presents the estimated impacts of the simulated self-efficacy intervention on STEM completion rates for policy 'always-takers', as defined in Section 6. The second panel 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 third panel presents the aggregate effects on STEM completion rates for women above the math ability median.

Impact on STEM Completion Rates. While increasing enrollment rates is an important first step for gauging the effectiveness of any STEM-promoting policy, this effect may not translate into increased graduation rates. Unlike the effect at enrollment, the simulated intervention may affect

math self-efficacy intervention could have larger impacts on STEM participation if it shaped women's preferences to increase interest and enjoyment from studying these fields.

STEM completion rates for 'always-takers' through the productivity of θ_{SE} in leading to successful STEM graduation. In the first panel of Figure 8, I show the impact of the intervention on always-takers' STEM completion rates across the θ_M distribution. Fitting in with the evidence presented in Section 4, I find that a self-efficacy boost for high math-ability women already enrolled in STEM would increase their completion rates, from 58.9% up to 62.3%. The second panel documents the corresponding impacts for 'compliers.' As these students were not originally enrolled in STEM, none of them would have completed a STEM major in absence of the intervention. However, while the self-efficacy boost would induce them to enroll in STEM, not all students would succesfully graduate from these majors. In fact, average STEM completion rates among 'compliers' would equal 62.1%. The combination of the impacts of the simulated policy on these two groups implies that average STEM completion rates for women above the θ_M median would increase from 3.8% to 4.6%, or 17% of baseline graduation rates (as shown in panel (c) of Figure 8). For women in the top decile of the math ability distribution, the policy would similarly boost completion rates by 1.1 percentage points, from 8.1% up to 9.2%.

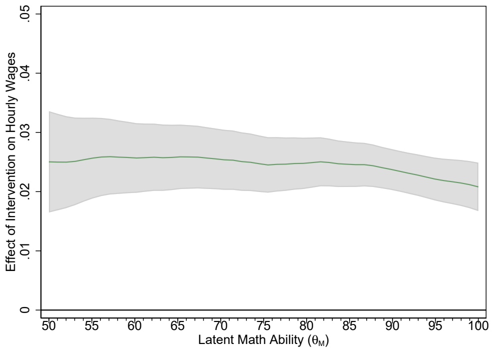
6.3 Impacts on Labor Market Outcomes

While I have so far shown that small self-efficacy-based interventions can lead to increased STEM participation rates for women, an open question remains about whether the simulated policy would lead to improved labor market outcomes, especially in the context of the heterogeneous returns to math-intensive majors shown in Section 5. A self-efficacy boost could affect labor market outcomes both through an increased likelihood of STEM completion but also through increased labor market productivity, given the positive direct impact of non-cognitive skills on labor market outcomes previously shown in the literature (Heckman et al., 2006).[44]

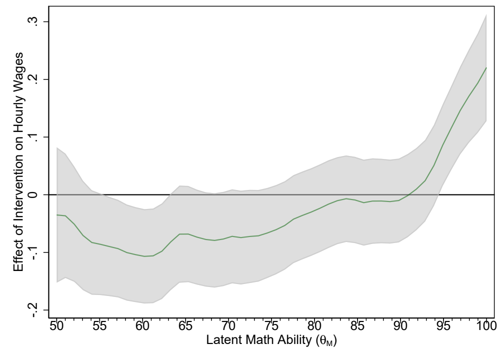
In Figure 9, I present the impact of the simulated policy intervention on students' hourly wages. The first panel considers impacts for always-takers. I find that the self-efficacy boost would increase average hourly wages by 2.4%, with largely constant effects across the θ_M distribution. The estimated impact can be explained both by the increased likelihood that 'always-takers' would complete a STEM degree — which has sizable wage returns (Figure 6), but also through the direct productivity of θ_{SE} in the labor market: a one σ increase in self-efficacy for students enrolled in STEM would increase hourly wages by 8.3%. Meanwhile, the second panel of Figure 9 presents the estimated effects for compliers, for whom the estimated impact depends both on the majors they are switching from and the heterogeneous returns to STEM enrollment vis-à-vis their original majors. In light of the returns to STEM majors presented in Table 2, I find that the policy simulation would lower wages for women closer to the θ_M median. On the other hand, given the large returns to STEM participation for women in the top math ability decile, high math ability 'compliers' would enjoy substantially higher wages (exceeding 20% at the top of the distribution) from the self-efficacy boost.

[44]Higher mathematical self-efficacy could lead students working in math-intensive occupations to gain confidence in their ability to perform well on math-related tasks, thus leading to increased productivity and higher wages.

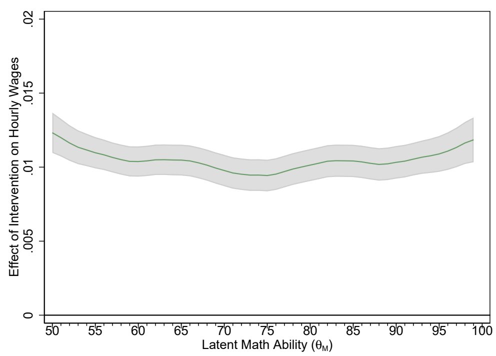
Figure 9: Estimted Impacts of Policy Intervention on Hourly Wages



(a) Impact on Hourly Wages: Always-Takers



(b) Impact on Hourly Wages: Compliers



(c) Impact on Hourly Wages: Overall

Note: The first panel of Figure 9 presents the estimated impacts of the simulated self-efficacy intervention on hourly wages for policy 'always-takers', as defined in Section 6. The second panel presents the estimated impact of the policy for 'compliers'. The third panel presents the aggregate effects of the policy on the early-career hourly wages of women above the θ_M median.

In Figure A.14, I show that the self-efficacy boost would still improve labor market outcomes for never-takers, as θ_{SE} still has a modest impact on labor marekt productivity for students in non-STEM majors. As such, the third panel of Figure 9 shows that the simulated policy intervention would boost average hourly wages by close to 1%, with largely homogeneous effects across the math ability distribution. Nonetheless, this effect masks sizably heterogeneous impacts for students depending on whether they change their majors in light of the self-efficacy increase and in particular, depending on which major they switch out of.

While the magnitude of the overall wage effect may not appear to be economically significant, the simulated self-efficacy intervention is small in magnitude. Larger labor market impacts could potentially be achieved with improved math-based non-cognitive skill development for women prior to college entry. Moreover, the proposed intervention would have a sizable impact on the labor market outcomes of high-achieving women who would switch into STEM, thus indicating that in a context of time/budget constraints, identifying individuals who would benefit the most from such interventions is of paramount importance. As such, policymakers could target various policy interventions towards high-math achievers. All in all, well-targeted skill development policies could thus increase women's participation in STEM while improving early-career labor market outcomes.

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 the existing literature has largely focused on the importance of test scores and preferences as drivers of major choices, the importance of non-cognitive skills has received scant attention in the 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 further shown a lower correlation between math ability and self-efficacy for women than for men, indicating a relative shortfall of high-skilled women who are confident in their math abilities. This finding is particularly 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 thus reduces their participation in STEM majors. As such, future research should further examine the drivers of gender differences in mathematical self-efficacy given their importance in shaping gender gaps in STEM.

Given the focus on increasing women's STEM participation rates, I have also analyzed whether women's labor market outcomes would improve from STEM participation. While increasing participation in these fields may bring important non-pecuniary benefits, I find significant heterogeneity in the wage returns in STEM, depending both on the alternative major under consideration but also across the math ability distribution. As I find large returns to STEM participation for high math ability women, STEM-promoting policies targeted towards high math-achievers would also deliver improved labor market outcomes. Lastly, following an extensive literature showing the malleability of non-cognitive skills through late adolescence, I have explored whether self-efficacy interventions could help in closing gender gaps in STEM. I have found that small skill development interventions could increase STEM enrollment and graduation rates by 10-15 percent relative to baseline levels, with larger impacts for high math performers. Furthermore, as the self-efficacy boost results in a small increase in hourly wages for top female math achievers, non-cognitive skill development interventions offer a promising pathway for future policy development.

References

Ahn, T., P. Arcidiacono, A. Hopson, and J. R. Thomas (2019). 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. *Annu. Rev. Econ. 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.

Attanasio, O., R. Blundell, G. Conti, and G. Mason (2020). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 104171.

Aucejo, E. M., J. James, et al. (2019). The path to college education: Are verbal skills more important than math skills? Technical report.

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

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

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bordón, P., C. Canals, and A. Mizala (2020). The gender gap in college major choice in Chile. *Economics of Education Review 77*, 102011.

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. Technical Report 2.

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.

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

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.

Fortin, N. M. (2008). The gender wage gap among young adults in the United States: The importance of money versus people. *Journal of Human Resources 43*(4), 884–918.

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

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.

Hamermesh, D. S. and S. G. Donald (2008). The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. *Journal of Econometrics 144*(2), 479–491.

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., 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.

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

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

Humphries, J. E., 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.

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

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.

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 (2017). Choice of majors: Are women really different from men? Technical report, National Bureau of Economic Research.

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, January). 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), 000–000.

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.

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

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*.

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.

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

Ransom, T. (2020). Selective migration, occupational choice, and the wage returns to college majors.

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. S. Urzúa (2018). Dynamic treatment effects of job training. Technical report, National Bureau of Economic Research.

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.

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*, 102033.

Sanders, C. (2015). Reading skills and earnings: Why do doing words good hurt you're wages. Technical report.

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 Appendix Tables and Figures

Table A.1: Sample Restrictions

Sample Restriction	Sample Size	Women	Men
Full Sample	16,197	8,107	8,090
Baseline Test Takers	15,892	7,975	7,917
Self-Efficacy Measures	13,446	6,876	6,570
Four-Year College Enrollment	5,183	2,899	2,284
Missing Educational Attainment	4,980	2,803	2,177
Missing Labor Market Outcomes	4,599	2,615	1,984

Source: Educational Longitudinal Study of 2002.
Note: Table A.1 shows the sample restrictions imposed on the baseline ELS sample result in the final sample used in the paper.

Table A.2: Determinants of Missing Test Scores

	BY Math	F1 Math	GPA	BY SE	F1 SE	Control	Motivation	Action
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Both Parents	0.001	0.012	-0.006	0.003	0.044**	0.008	0.006	0.003
	(0.002)	(0.011)	(0.012)	(0.019)	(0.019)	(0.020)	(0.019)	(0.020)
Both Parents × Male	-0.001	0.022	0.025	0.006	-0.035	-0.007	0.007	0.018
	(0.002)	(0.017)	(0.018)	(0.032)	(0.028)	(0.032)	(0.032)	(0.033)
Parental Education	0.000	0.001	-0.002	0.007**	-0.002	0.006**	0.007**	0.008***
	(0.000)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Parental Education× Male	-0.000	0.001	0.001	-0.008**	0.005*	-0.007*	-0.008**	-0.007**
	(0.000)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
HH Income	-0.000	0.004**	0.007**	0.001	0.007**	0.004	0.005	0.004
	(0.000)	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
HH Income × Male	0.000	-0.003	-0.002	0.007	-0.003	0.006	0.006	0.004
	(0.000)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Constant	0.997***	0.897***	0.900***	0.694***	0.767***	0.666***	0.644***	0.643***
	(0.004)	(0.024)	(0.030)	(0.055)	(0.047)	(0.055)	(0.055)	(0.056)
Observations				4,599				
R^2	0.001	0.005	0.004	0.006	0.004	0.005	0.006	0.006

Note: Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.2 presents evidence from a linear regression on the determinants of missing test scores and observed self-efficacy and non-cognitive skill measures.

Table A.3: 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	0.012	-0.017	0.009
	(0.011)	(0.014)	(0.016)	(0.018)
Parental Education	0.004*	0.002	-0.006*	-0.006*
	(0.002)	(0.002)	(0.003)	(0.003)
Log HH Income	0.003	0.002	-0.005	0.003
	(0.002)	(0.002)	(0.003)	(0.004)
Underrepresented Minority	0.027*	0.027*	-0.016	-0.008
	(0.011)	(0.013)	(0.018)	(0.019)
Baseline Math Exam	0.017*	0.011	-0.012	-0.027***
	(0.007)	(0.006)	(0.008)	(0.008)
HS GPA	-0.000	0.016*	-0.001	0.001
	(0.006)	(0.007)	(0.008)	(0.008)
Baseline Math Self-Efficacy	0.017***	0.008	0.028***	0.011
	(0.005)	(0.006)	(0.008)	(0.009)
Non-Cognitive Skills (PCA)	-0.003	0.013	-0.019*	-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	-0.017	-0.013	0.056*	0.013
	(0.023)	(0.014)	(0.026)	(0.013)
Parental Education	-0.002	0.005	0.001	-0.004*
	(0.004)	(0.003)	(0.004)	(0.002)
Log HH Income	0.006	-0.005*	-0.006	-0.003
	(0.005)	(0.002)	(0.004)	(0.002)
Underrepresented Minority	0.076**	0.026	0.005	-0.009
	(0.024)	(0.015)	(0.025)	(0.013)
Baseline Math Exam	0.043***	0.011	-0.016	-0.016***
	(0.011)	(0.006)	(0.009)	(0.005)
HS GPA	0.037***	0.020**	-0.006	0.012*
	(0.010)	(0.007)	(0.009)	(0.005)
Baseline Math Self-Efficacy	0.074***	-0.000	-0.010	-0.004
	(0.013)	(0.008)	(0.013)	(0.005)
Non-Cognitive Skills (PCA)	-0.035**	0.005	0.009	0.009
	(0.013)	(0.007)	(0.012)	(0.005)

Note: Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.3 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.

Table A.4: 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	0.012	-0.016	0.009
	(0.011)	(0.014)	(0.016)	(0.018)
Parental Education	0.005*	0.002	-0.004	-0.005
	(0.002)	(0.002)	(0.003)	(0.003)
Log HH Income	0.003	0.002	-0.004	0.004
	(0.002)	(0.002)	(0.003)	(0.004)
Underrepresented Minority	0.026*	0.027*	-0.020	-0.011
	(0.010)	(0.013)	(0.018)	(0.019)
Baseline Math Exam	0.020**	0.013	0.013	-0.010
	(0.007)	(0.008)	(0.009)	(0.009)
HS GPA	0.000	0.016*	0.002	0.003
	(0.006)	(0.008)	(0.008)	(0.008)
Baseline Math Self-Efficacy	0.016**	0.008	0.021*	0.006
	(0.005)	(0.006)	(0.008)	(0.010)
Non-Cognitive Skills (PCA)	-0.002	0.013*	-0.015	0.002
	(0.005)	(0.007)	(0.008)	(0.010)
Baseline Reading Exam	-0.007	-0.004	-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	-0.013	0.057*	0.013
	(0.023)	(0.014)	(0.025)	(0.013)
Parental Education	-0.000	0.005	0.002	-0.004*
	(0.004)	(0.003)	(0.004)	(0.002)
Log HH Income	0.005	-0.005*	-0.007	-0.003
	(0.005)	(0.002)	(0.004)	(0.002)
Underrepresented Minority	0.073**	0.026	0.004	-0.009
	(0.024)	(0.015)	(0.025)	(0.013)
Baseline Math Exam	0.069***	0.011	0.002	-0.015*
	(0.013)	(0.007)	(0.012)	(0.006)
HS GPA	0.041***	0.020**	-0.004	0.012*
	(0.010)	(0.007)	(0.009)	(0.005)
Baseline Math Self-Efficacy	0.071***	-0.001	-0.015	-0.004
	(0.013)	(0.008)	(0.013)	(0.005)
Non-Cognitive Skills (PCA)	-0.032*	0.005	0.012	0.009
	(0.013)	(0.007)	(0.012)	(0.005)
Baseline Reading Exam	-0.043***	0.000	-0.029**	-0.002
	(0.011)	(0.006)	(0.011)	(0.005)

Note: Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 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. It follows the results presented in Table A.3 while including baseline reading test scores as a control variable. The omitted category is enrollment in a major in the 'Other' category.

Table A.5: Determinants of STEM Completion

	Women	Men
	(1)	(2)
Both Parents	0.124	0.031
	(0.102)	(0.065)
Parental Education	-0.010	-0.000
	(0.018)	(0.011)
Log HH Income	-0.035	0.007
	(0.022)	(0.015)
Underrepresented Minority	-0.110	0.028
	(0.116)	(0.073)
Baseline Math Exam	0.094*	0.090***
	(0.051)	(0.029)
HS GPA	0.129***	0.139***
	(0.046)	(0.028)
Baseline Math Self-Efficacy	0.101	-0.005
	(0.067)	(0.043)
Non-Cognitive Skills (PCA)	-0.086	0.006
	(0.066)	(0.033)
Observations	119	369
R^2	0.231	0.155

Note: Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The first and second columns of Table A.5 presents evidence from a linear regression examining the determinants of STEM completion for female and male students initially enrolled in STEM, respectively.

Table A.6: OLS Estimates: Wage Returns to Majors by Gender

	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)
STEM	0.067	0.027	0.028	0.268***	0.249***	0.243***
	(0.063)	(0.062)	(0.062)	(0.035)	(0.036)	(0.036)
Life Science	-0.042	-0.069	-0.069	0.006	-0.012	-0.014
	(0.047)	(0.048)	(0.048)	(0.066)	(0.066)	(0.066)
Business	0.108***	0.104***	0.106***	0.178***	0.171***	0.167***
	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.033)
Health	0.258***	0.261***	0.262***	0.321***	0.307***	0.303***
	(0.034)	(0.033)	(0.033)	(0.084)	(0.084)	(0.083)
Background Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Test Scores	No	Yes	Yes	No	Yes	Yes
Reading Test	No	No	Yes	No	No	Yes
Observations		2,013			1,524	

Note: Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table A.6 presents evidence from a linear wage regression examining the returns to enrolling in different majors for women (first three columns) and for men (last three columns). The first and fourth columns include individual and family background characteristics as control variables. The second and fifth columns add baseline test scores as controls. The third and sixth columns further add baseline reading test scores as controls. The sample used in the analysis includes individuals in the estimation sample (see Table A.1) who worked at least 500 hours in 2011.

Table A.7: College Major Choices and Missing Test Scores

	Women				Men			
	STEM	Life Sciences	Health	Other	STEM	Life Sciences	Health	Other
Missing Test Score	-0.002	-0.008	0.003	-0.000	0.025	-0.011	-0.030	-0.008
	(0.008)	(0.010)	(0.013)	(0.014)	(0.018)	(0.011)	(0.017)	(0.009)
Observations	2,615				1,984			

Note: Robust standard errors in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Table A.7 presents the estimated marginal effects from a multinomial logit examining the determinants of initial major choices for women (first four columns) and men (last four columns). The omitted category is enrollment in a major in the 'Other' category.

Table A.8: Variables Used in Implementation of the Model

	Math/SE Measures (1)	NC Measures (2)	Initial Major (3)	Dropout (4)	Final Major (5)	Employment (6)	Wage Eq. (7)
Observables							
Race	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Composition	Yes	Yes	Yes	Yes	Yes		
Parents' Education	Yes	Yes	Yes	Yes	Yes		
Family Income	Yes	Yes	Yes	Yes	Yes		
Parents' Math Measures	Yes						
Latent Ability							
Math Ability	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-Cognitive Ability	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Math Self-Efficacy	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Table A.8 presents the variables used in the empirical model. In the measurement system, I use two math test scores, students' high school GPA, two measures of mathematical self-efficacy and three measures of baseline non-cognitive skills. Initial major choices are defined in five categories: math-intensive STEM, life sciences, business, health-related fields and other majors. The dropout decision encompasses continuing in four-year college or not. The final major decision involves staying in the original major or switching to a different field. Hourly wages are measured in logs at age 25.

Table A.9: Estimated Returns to STEM v. Alternative Majors: Model and Reduced Form Estimates

Panel A. Women				
Estimate	Life Sciences	Business	Health	Other
ATE	0.072	-0.117	-0.250	-0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)
OLS (X)	0.112	-0.030	-0.178	0.064
	(0.076)	(0.060)***	(0.065)***	(0.052)
OLS (θ)	0.100	-0.064	-0.256	0.024
	(0.078)	(0.061)***	(0.067)***	(0.052)***
NNM-3	0.099	-0.076	-0.248	0.022
	(0.080)	(0.068)	(0.074)***	(0.069)

Panel B. Men				
Estimate	Life Sciences	Business	Health	Other
ATE	0.254	0.060	-0.167	0.217
	(0.001)***	(0.001)***	(0.003)***	(0.001) ***
OLS (X)	0.245	0.091	-0.072	0.268
	(0.069)***	(0.041)***	(0.083)***	(0.034)***
OLS (θ)	0.243	0.076	-0.100	0.246
	(0.070)***	(0.044)***	(0.087)***	(0.035)***
NNM-3	0.241	0.065	-0.180	0.202
	(0.076)***	(0.046)	(0.098)**	(0.039)***

Note: Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Table A.9 presents the estimated returns to STEM enrollment vis-à-vis other majors for both females (Panel A) and males (Panel B). I compare the estimated model-based average treatment effect to those recovered under alternative estimation strategies, including OLS with background characteristics as controls, OLS with background characteristics and test scores as controls, and a nearest-neighbor matching estimator.

Table A.10: Decomposition of Estimated Returns to STEM v. Alternative Majors by Gender

		Panel A. Women		
Estimate	Life Sciences	Business	Health	Other
Observed Differences	0.095	-0.044	-0.173	0.049
	(0.005)***	(0.004)***	(0.004)***	(0.003)***
Pairwise ATE	0.078	-0.074	-0.240	-0.008
	(0.004)***	(0.003)***	(0.003)***	(0.001)***
Selection Bias	0.010	0.059	0.096	0.059
	(0.005)**	(0.004)***	(0.004)***	(0.003)***
Sorting Gains	0.007	-0.028	-0.029	-0.003
	(0.007)	(0.006)***	(0.006)***	(0.005)
		Panel B. Men		
Estimate	Life Sciences	Business	Health	Other
Observed Differences	0.243	0.082	-0.099	0.250
	(0.004)***	(0.003)***	(0.009)***	(0.002)***
Pairwise ATE	0.279	0.076	-0.190	0.220
	(0.003)***	(0.002)***	(0.006)***	(0.001)***
Selection Bias	-0.048	-0.021	0.097	0.021
	(0.005)***	(0.003)***	(0.015)***	(0.002)***
Sorting Gains	0.013	0.027	-0.006	0.009
	(0.004)***	(0.003)***	(0.008)	(0.003)***

Note: Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Table A.10 presents the decomposition introduced in () 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.

Table A.11: 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	—	10.3	8.5	8.0	13.6
Life Sciences	14.0	—	11.9	12.2	19.8
Business	16.6	16.6	—	19.3	30.0
Health	18.0	19.6	22.1	—	36.6
Other	51.4	53.6	57.5	60.5	—
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	—	25.4	23.8	20.3	34.6
Life Sciences	13.7	—	11.3	10.4	16.2
Business	24.3	21.2	—	22.4	37.8
Health	7.0	6.8	8.0	—	11.5
Other	55.0	46.6	56.8	46.9	—
Total	100	100	100	100	100

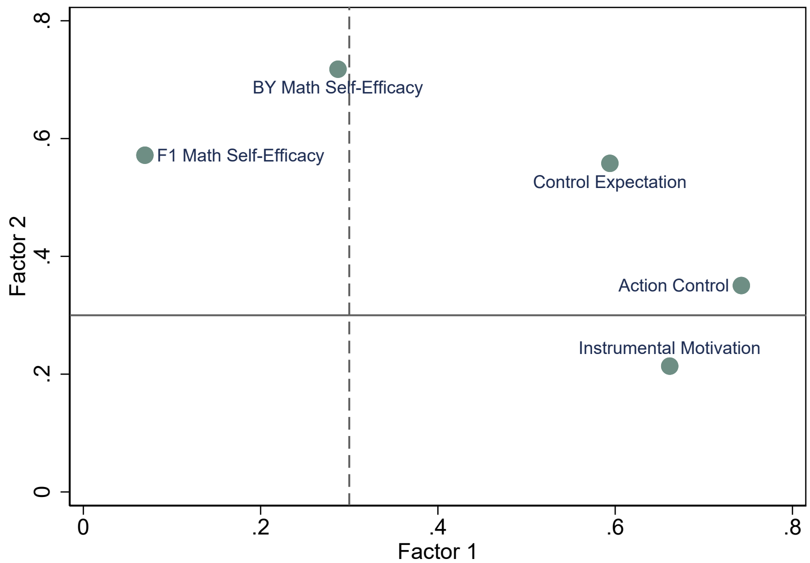
Note: For individuals enrolled in different majors (presented in different columns), Table A.11 documents the share of students with different second best options. The first panel presents evidence for women enrolled across the five major groupings in the model (STEM, life sciences, business, health and 'Other') and the second panel shows evidence for men.

Table A.12: Conditional Returns to Initial Major Completion by Gender

	Panel A. Women				
Estimate	STEM	Life Sciences	Business	Health	Other
	(1)	(2)	(3)	(4)	(5)
ATE (v. Switching)	0.296	-0.042	-0.080	0.283	-0.109
	(0.007)***	(0.004)***	(0.003)***	(0.003)***	(0.001)***
ATE (v. Dropout)	0.257	0.031	0.156	0.371	0.179
	(0.006)***	(0.004)***	(0.003)***	(0.003)***	(0.001)***
	Panel B. Men				
Estimate	STEM	Life Sciences	Business	Health	Other
	(1)	(2)	(3)	(4)	(5)
ATE (v. Switching)	0.356	-0.169	0.070	-0.086	-0.126
	(0.003)***	(0.007)***	(0.003)***	(0.010)***	(0.001)***
ATE (v. Dropout)	0.564	-0.045	0.147	-0.698	0.119
	(0.002)***	(0.006)***	(0.003)***	(0.047)***	(0.002)***

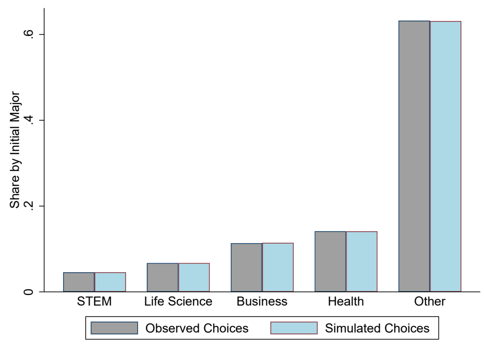
Note: Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Table A.12 presents evidence on the conditional returns to STEM completion relative to degree switching (equation (13)) and relative to college dropout (equation (14)). The first panel presents evidence for women, whereas the second panel presents evidence for men.

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

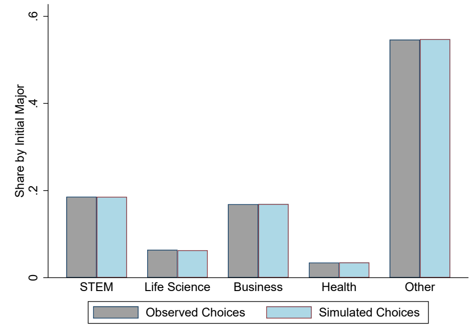


Note: Figure A.1 presents the estimated loadings from exploratory factor analysis of the three non-cognitive skill measures available in the ELS along with the baseline and follow-up math self-efficacy measres with orthogonal factors. The solid horizontal and dashed vertical lines are placed at 0.30,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

Figure A.2: Goodness of Fit: Initial Majors



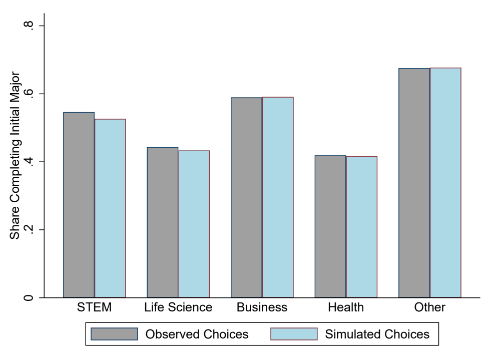
(a) Initial Major Choices: Women



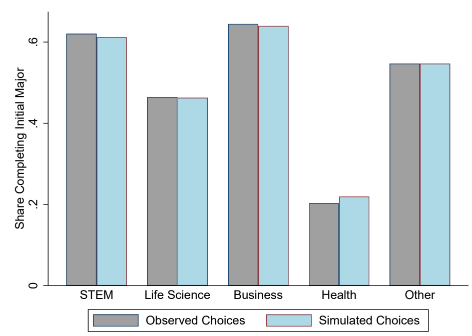
(b) Initial Major Choices: Men

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



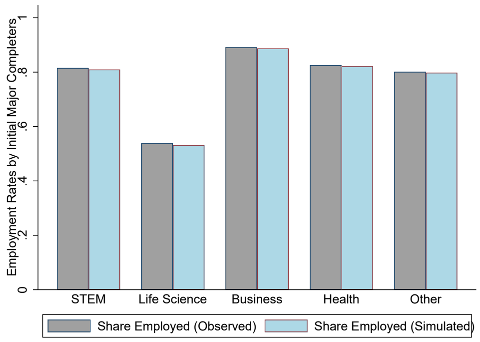
(a) % Completing Initial Major: Women



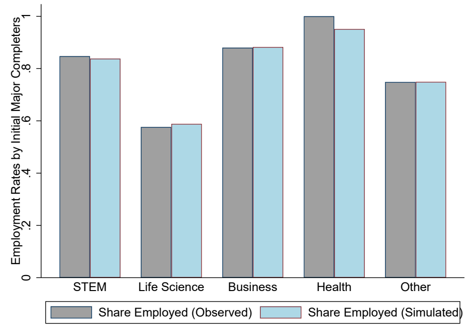
(b) % Completing Initial Major: Men

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 before 2012 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



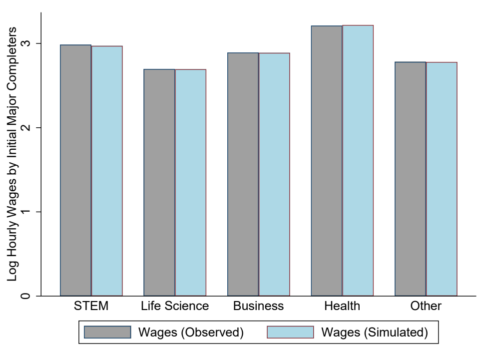
(a) % Employed by Major: Women



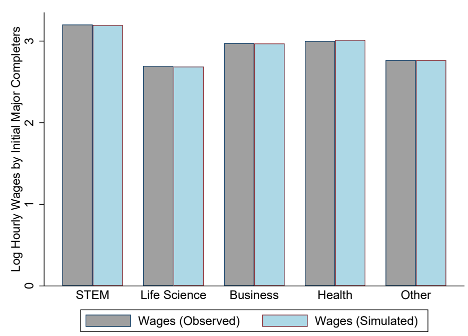
(b) % Employed by Major: Men

Note: Figure A.4 presents the observed proportion of students who are employed in 2011 among those who completed their initial major and compares it to the model-based simulated share of initial-major completers who are employed. The first panel presents evidence for women and the second panel presents evidence for men.

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



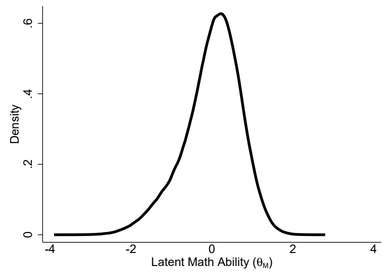
(a) Wages by Major: Women



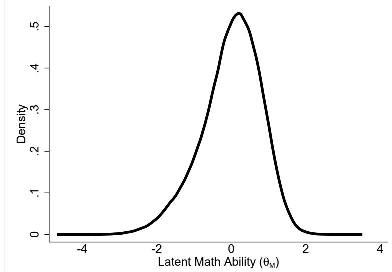
(b) Wages by Major: Men

Note: The first panel of Figure A.5 presents observed average hourly wages among those who completed their initial major and were employed in 2011, which is then compared to model-based average wages for initial major completers who are employed. The first panel presents evidence for women and the second panel presents evidence for men.

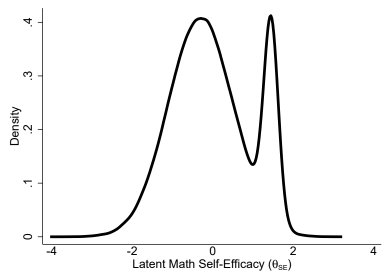
Figure A.6: Density of Latent Ability by Gender



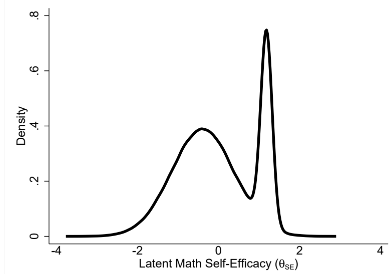
(a) Density of θ_M: Women



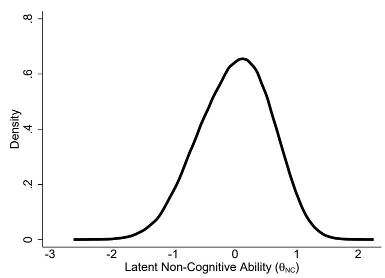
(b) Density of θ_M: Men



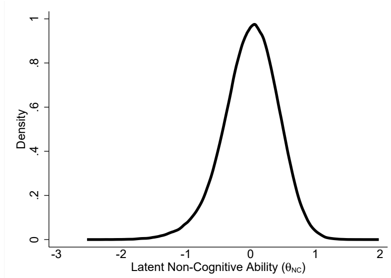
(c) Density of θ_{SE}: Women



(d) Density of θ_{SE}: Men



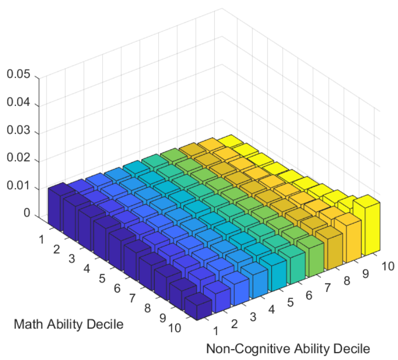
(e) Density of θ_{NC}: Women



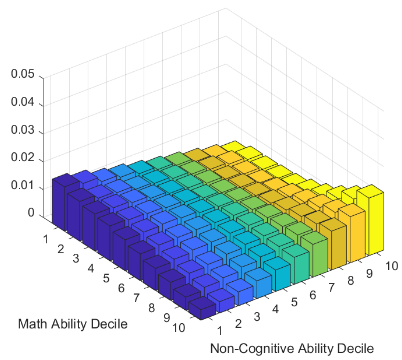
(f) Density of θ_{NC}: Men

Note: Figure A.6 shows the marginal densities of the estimated latent math ability, math self-efficacy and non-cognitive skill factors. Panels (a), (c) and (e) present the estimated marginal densities for women, whereas panels (b), (d) and (f) present corresponding evidence for men.

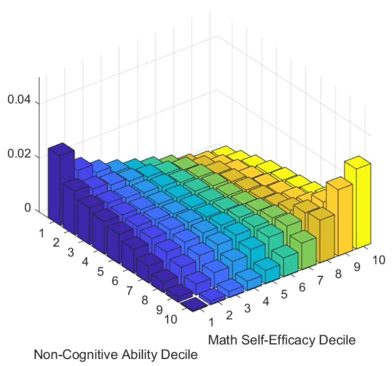
Figure A.7: Correlation of Latent Factors by Gender



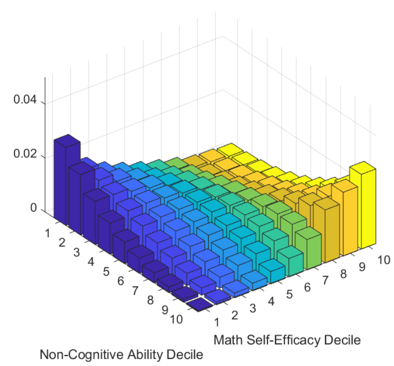
(a) Joint Distribution of θ_M and θ_{NC}: Women



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



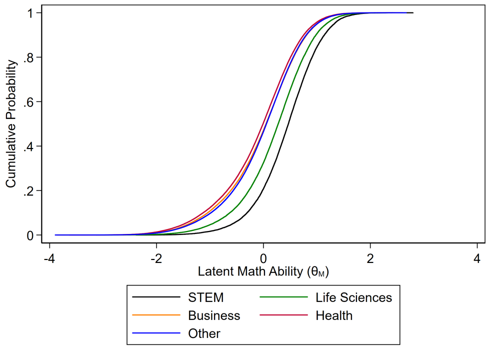
(c) Joint Distribution of θ_{NC} and θ_{SE}: Women



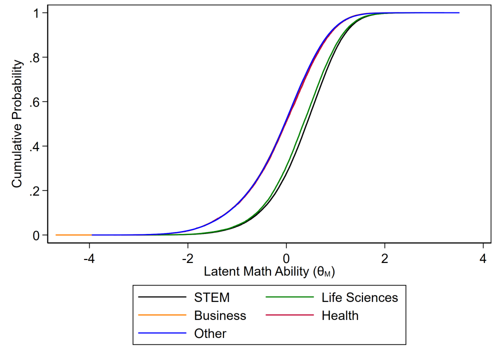
(d) Joint Distribution of θ_{NC} and θ_{SE}: Men

Note: The first two panels of Figure A.7 present the joint density of math ability and non-cognitive by gender, documenting the share of individuals pertaining to the joint decile of the two latent factor distributions. The first panel presents results for women and the second panel presents evidence for men. The last two panels present corresponding evidence for the joint distribution of mathematical self-efficacy and non-cognitive ability, with panel (c) showing evidence for women and panel (d) presenting evidence for men.

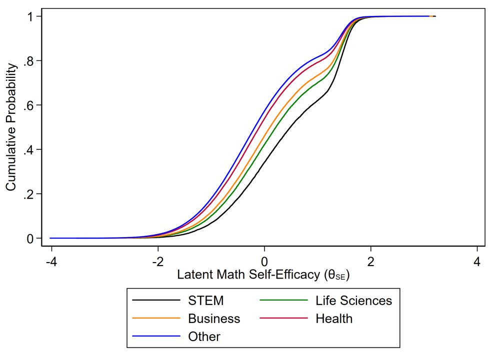
Figure A.8: Initial Major Choices by Gender



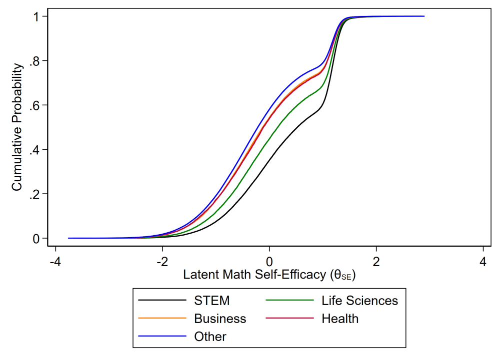
(a) θ_M by Major: Women



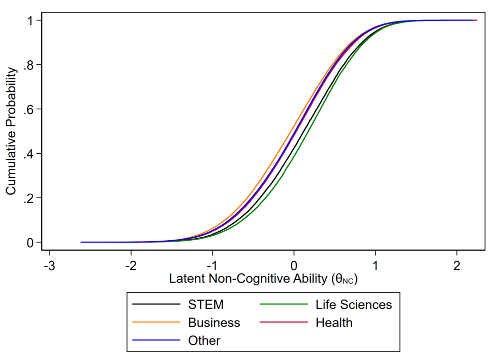
(b) θ_M by Major: Men



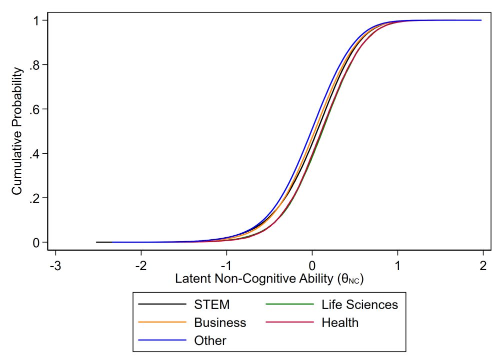
(c) θ_{SE} by Major: Women



(d) θ_{SE} by Major: Men



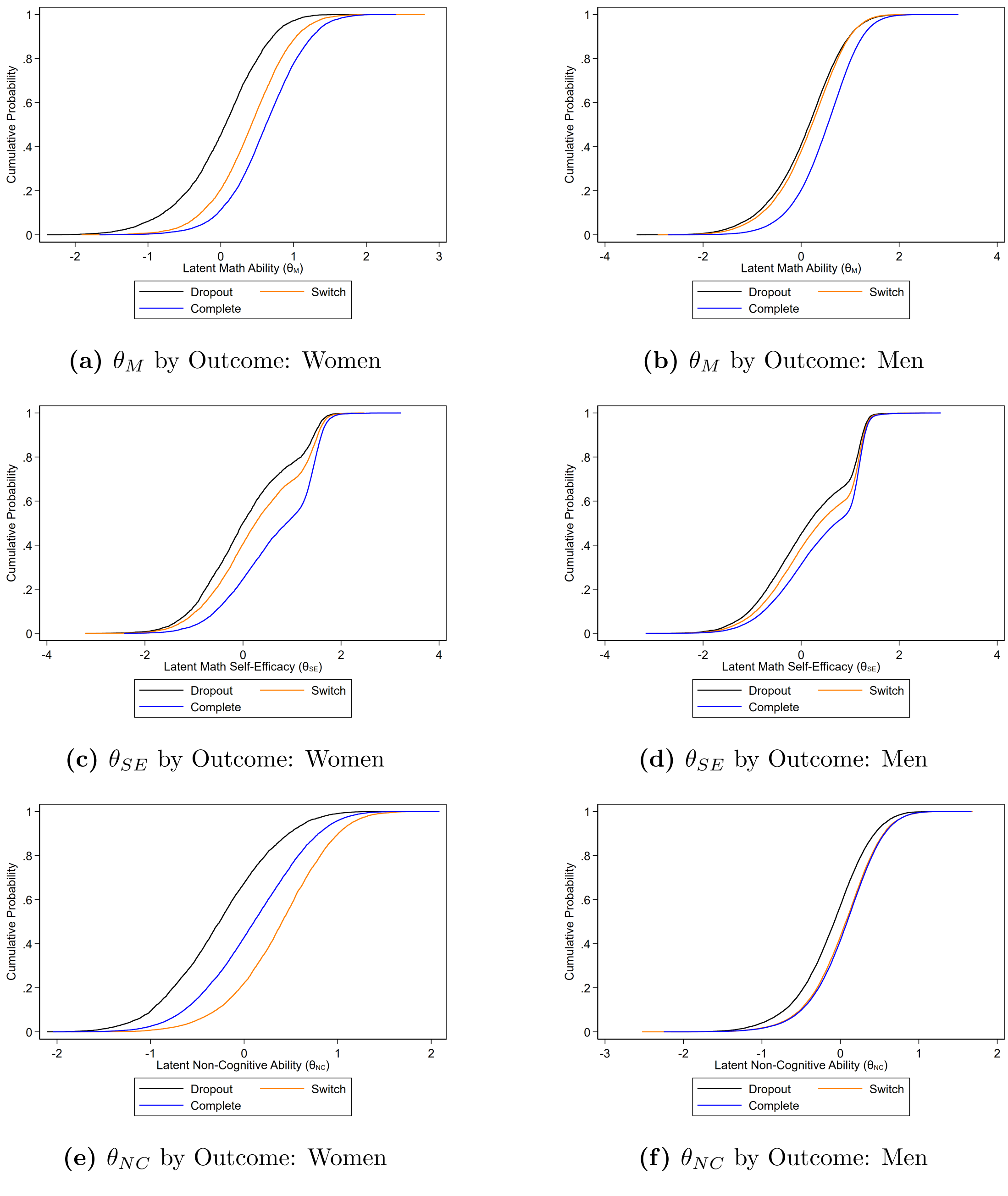
(e) θ_{NC} by Major: Women



(f) θ_{NC} by Major: Men

Note: Figure A.8 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.9: 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

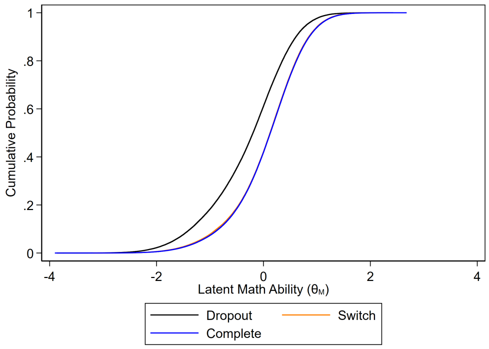
(d) θ_{SE} by Outcome: Men

(e) θ_{NC} by Outcome: Women

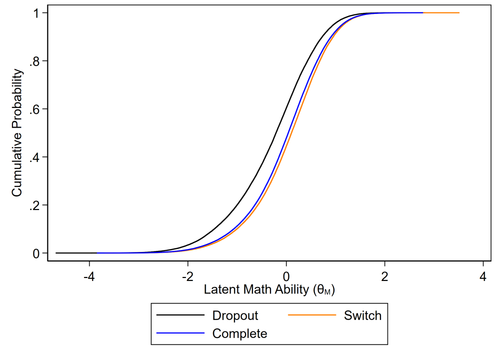
(f) θ_{NC} by Outcome: Men

Note: Figure A.9 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.

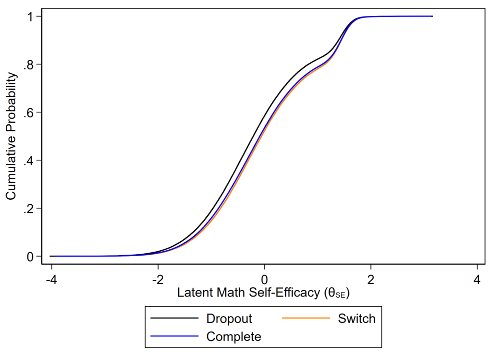
Figure A.10: Graduation Outcomes Among Initial Non-STEM Enrollees by Gender



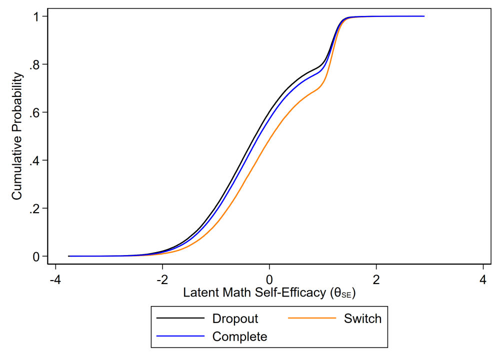
(a) θ_M by Outcome: Women



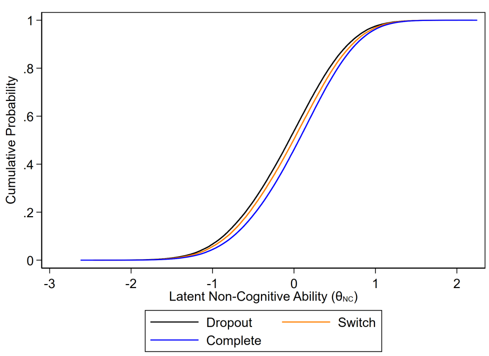
(b) θ_M by Outcome: Men



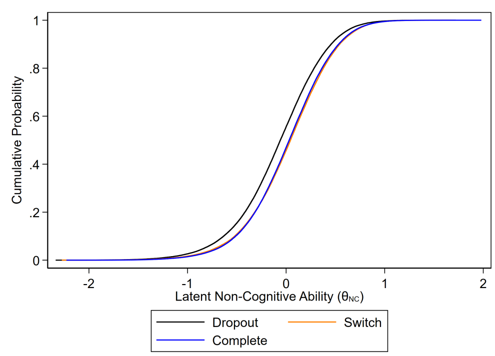
(c) θ_{SE} by Outcome: Women



(d) θ_{SE} by Outcome: Men



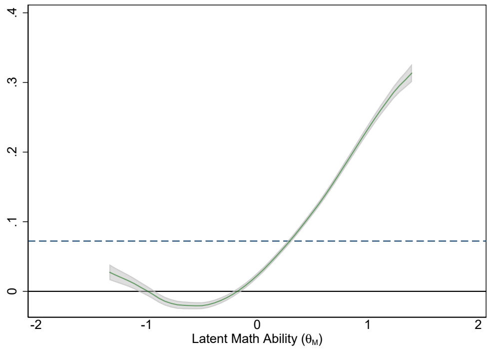
(e) θ_{NC} by Outcome: Women



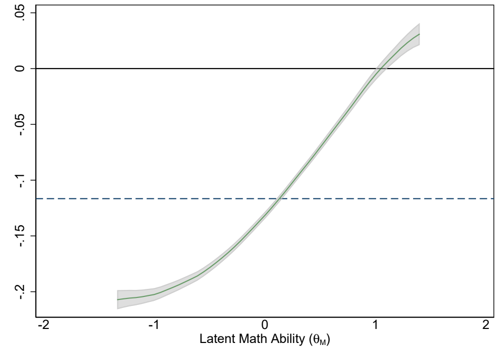
(f) θ_{NC} by Outcome: Men

Note: Figure A.10 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.

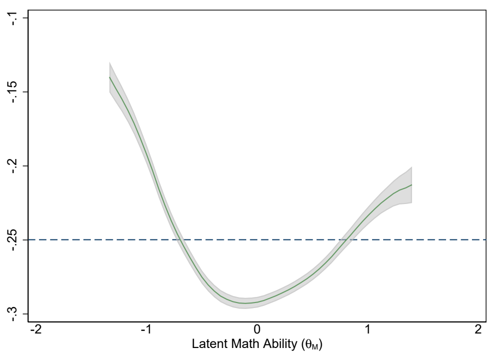
Figure A.11: Heterogeneous Returns to STEM Enrollment Across θ_M Distribution for Women



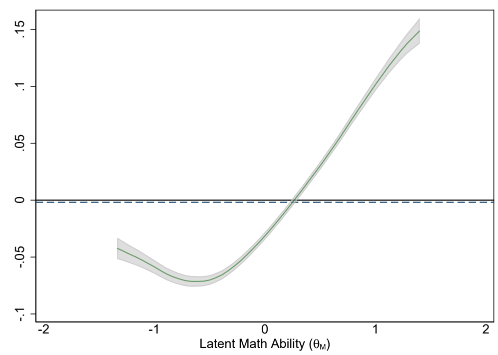
(a) v. Life Sciences



(b) v. Business



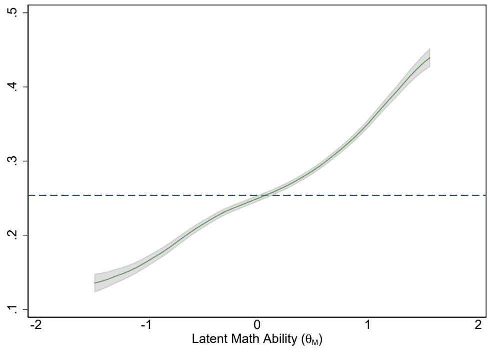
(c) v. Health



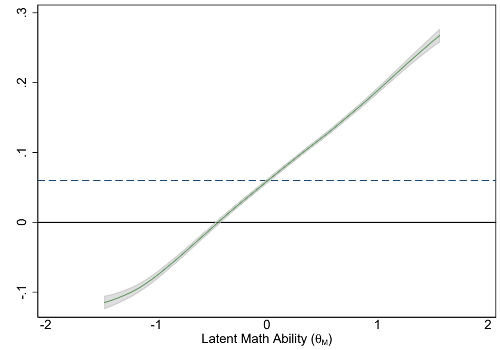
(d) v. Other

Note: Figure A.11 presents heterogeneous returns to STEM enrollment relative to life sciences, business, health and 'Other' majors for women across the latent math ability (θ_M^f) distribution.

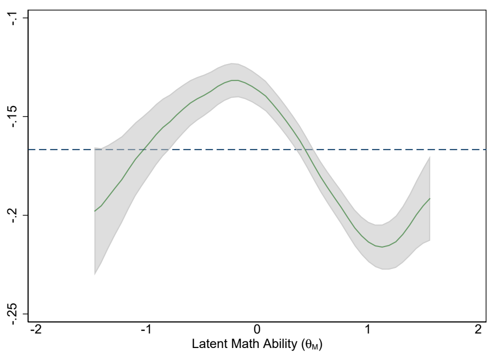
Figure A.12: Heterogeneous Returns to STEM Enrollment Across θ_M Distribution for Men



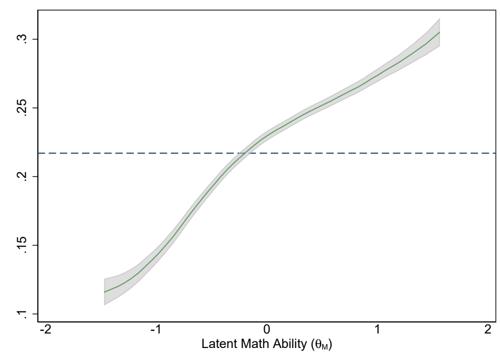
(a) v. Life Sciences



(b) v. Business



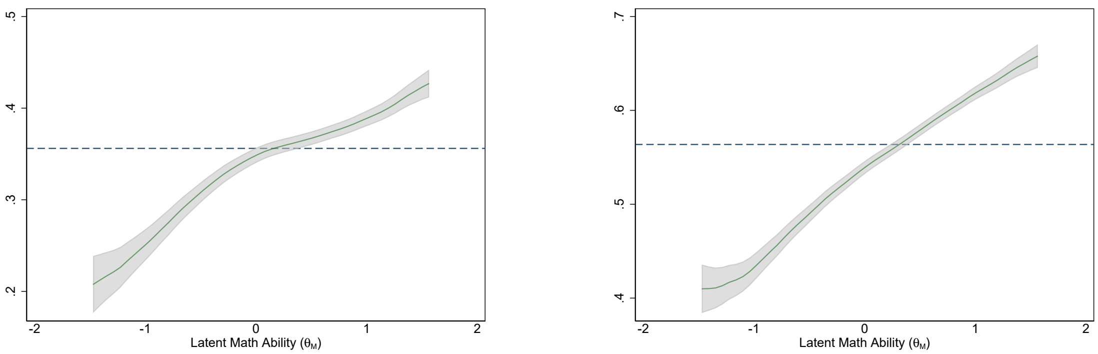
(c) v. Health



(d) v. Other

Note: Figure A.12 presents heterogeneous returns to STEM enrollment relative to life sciences, business, health and 'Other' majors for men across the latent math ability (θ_M^m) distribution.

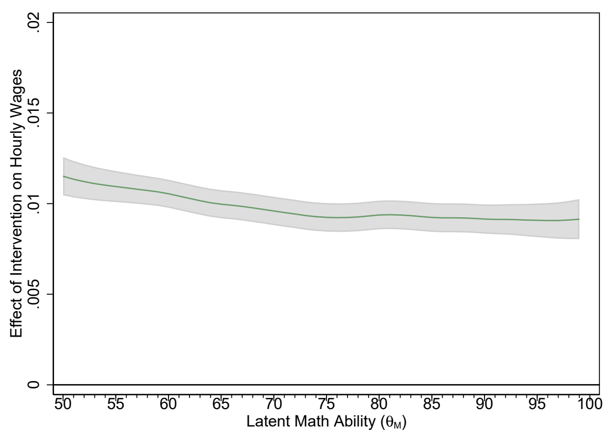
Figure A.13: Heterogeneous Returns to STEM Completion for Male STEM Enrollees



(a) Returns v. Switching **(b)** Returns v. Dropout

Note: The first panel of Figure A.13 presents heterogeneous returns to STEM completion relative to switching to a different major (equation (13)) for men. The second panel presents corresponding evidence relative to dropping out of college altogether (equation (14).

Figure A.14: Impacts of Intervention on Hourly Wages: Never-Takers



Note: Figure A.14 presents the estimated impacts of the simulated self-efficacy intervention on hourly wages for policy 'never-takers', as defined in Section 6.

B Identification of the Measurement System

This section presents the identification of the measurement system presented in Section 3. The identification of the distribution of unobserved ability follows the formal arguments presented in Carneiro et al. (2003); Hansen et al. (2004); 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:

$$Cov(SE_1, SE_2) = \alpha_1^{SE}\alpha_2^{SE}\sigma^2_{\theta_{SE}} \tag{B.16}$$

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:

$$Cov(SE_j, NC_k) = \alpha_j^{SE}\alpha_k^{SE}\sigma^2_{\theta_{SE}} + \alpha_j^{SE}\gamma_k^{NC}\sigma_{\theta_{SE},\theta_{NC}} \tag{B.17}$$

where $\sigma_{\theta_{SE},\theta_{NC}}$ captures the covariance of the latent math self-efficacy and non-cognitive skill factor. Equations (B.16)-(B.17) 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 ($\alpha_1^{SE} = 1$ and $\gamma_1^{NC} = 1$).[45] 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:

$$Cov(NC_k, NC_{k'}) = \alpha_k^{SE}\alpha_{k'}^{SE}\sigma^2_{\theta_{SE}} + \gamma_k^{NC}\gamma_{k'}^{NC}\sigma^2_{\theta_{NC}} + (\alpha_k^{SE}\gamma_{k'}^{NC} + \alpha_{k'}^{SE}\gamma_k^{NC})\sigma_{\theta_{SE},\theta_{NC}} \tag{B.18}$$

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 (B.16)-(B.17), $\sigma^2_{\theta_{NC}}$ 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:

$$\begin{aligned} Cov(M_l, M_{l'}) \;&=\; \alpha_l^{SE}\alpha_{l'}^{SE}\sigma^2_{\theta_{SE}} + \gamma_l^{NC}\gamma_{l'}^{NC}\sigma^2_{\theta_{NC}} + \eta_l^M\eta_{l'}^M\sigma^2_{\theta_M} + (\alpha_l^{SE}\gamma_{l'}^{NC} + \alpha_{l'}^{SE}\gamma_l^{NC})\sigma_{\theta_{SE},\theta_{NC}} \\ &\quad + (\alpha_l^{SE}\eta_{l'}^M + \alpha_{l'}^{SE}\eta_l^M)\sigma_{\theta_{SE},\theta_M} + (\gamma_l^{NC}\eta_{l'}^M + \gamma_{l'}^{NC}\eta_l^M)\sigma_{\theta_{NC},\theta_M} \end{aligned} \tag{B.19}$$

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

[45] 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:

$$\begin{aligned} Cov(M_l, NC_k) &= \alpha_l^{SE}\alpha_k^{SE}\sigma^2_{\theta_{SE}} + \gamma_l^{NC}\gamma_k^{NC}\sigma^2_{\theta_{NC}} + (\alpha_l^{SE}\gamma_k^{NC} + \alpha_k^{SE}\gamma_l^{NC})\sigma_{\theta_{SE},\theta_{NC}} \\ &+ (\alpha_k^{SE}\eta_l^M)\sigma_{\theta_{SE},\theta_M} + (\gamma_k^{NC}\eta_l^M)\sigma_{\theta_{NC},\theta_M} \end{aligned} \tag{B.20}$$

As such, the twelve covariances defined in equations (B.19)-(B.20) allow me to recover the remaining factor loadings (I normalize $\eta_1^M = 1$ for the baseline math test score), the variance of $theta_M$ ($\sigma^2_{\theta_M}$) and the two remaining covariances of the latent factors.[46] 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.[47]

[46] As discussed in Section 3.2, I additionally normalize the self-efficacy loading of high school GPA to zero ($\eta^M_{M3} = 0$).

[47] Freyberger (2018) extends the identification arguments presented by Kotlarski (1967) to a context with correlated factors.

C 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:

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

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

The block structure associated with the likelihood function is thus:

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

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

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

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

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

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

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

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

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,

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

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

Following the arguments presented in (C.2):

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

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:

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

4. Precisions:

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

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

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

Lastly, the posterior for θ becomes:

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

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:

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

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,

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

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:

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

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

Iterate over m until convergence.

D Estimated Model Parameters

Table D.1: Measurement System Loadings: Women

	BY SE	F1 SE	BY Math	F1 Math	GPA	Control	Motivation	Action
Constant	-0.24	-0.71	-0.96	-1.11	-0.24	-0.13	0.27	0.14
	(0.11)	(0.17)	(0.12)	(0.12)	(0.12)	(0.12)	(0.14)	(0.13)
Both Parents	0.08	0.09	0.16	0.16	0.18	0.11	0.03	0.09
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Log HH Income	-0.01	-0.00	0.01	0.00	-0.00	-0.01	-0.01	-0.02
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Parental Education	0.01	0.02	0.06	0.07	0.02	0.01	-0.02	0.00
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Underrepresented Minority	0.02	-0.03	-0.62	-0.56	-0.48	0.08	0.02	0.10
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Parent: Can Learn Math	-0.05	0.04	-0.05	-0.03				
	(0.02)	(0.02)	(0.02)	(0.02)				
Parent: Born Math Ability	0.00	0.06	-0.03	-0.03				
	(0.02)	(0.02)	(0.02)	(0.02)				
θ_{SE}	1.00	0.52	0.27	0.31	0.00	0.58	0.00	0.45
	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
θ_{NC}	0.00	0.00	-0.04	-0.04	0.26	1.00	1.14	1.44
	(0.00)	(0.00)	(0.03)	(0.03)	(0.03)	(0.00)	(0.03)	(0.04)
θ_M	0.00	0.00	1.00	1.21	0.60	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)
Sample Size				2,615				

Note: Standard errors in parenthesis. Table D.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.

Table D.2: Measurement System Loadings: Men

	BY SE	F1 SE	BY Math	F1 Math	GPA	Control	Motivation	Action
Constant	0.08	-0.00	-1.04	-0.58	-0.53	-0.21	-0.38	-0.04
	(0.10)	(0.19)	(0.15)	(0.14)	(0.14)	(0.13)	(0.16)	(0.15)
Both Parents	0.01	0.01	0.12	0.09	0.11	0.06	0.15	0.09
	(0.02)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Log HH Income	0.00	-0.01	0.00	-0.00	0.01	-0.01	0.00	-0.03
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Parental Education	0.01	0.03	0.07	0.06	0.02	0.02	0.01	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Underrepresented Minority	-0.02	-0.09	-0.54	-0.54	-0.58	-0.01	0.06	-0.01
	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Parent: Can Learn Math	0.01	-0.02	-0.01	-0.04				
	(0.01)	(0.03)	(0.02)	(0.02)				
Parent: Born Math Ability	-0.01	-0.03	0.03	-0.02				
	(0.01)	(0.02)	(0.02)	(0.02)				
θ_{SE}	1.00	0.53	0.33	0.33	0.00	0.74	0.00	0.55
	(0.00)	(0.02)	(0.02)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
θ_{NC}	0.00	0.00	-0.20	-0.21	0.39	1.00	1.77	2.25
	(0.00)	(0.00)	(0.05)	(0.04)	(0.05)	(0.00)	(0.07)	(0.08)
θ_M	0.00	0.00	1.00	1.17	0.51	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)
Sample Size				1,984				

Note: Standard errors in parenthesis. Table D.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.