Abilities and Career Choices: Addressing the Gender Gap in STEM Fields

Andreu Arenas*
Lídia Farré†
M.Cecilia Ramírez-Michelena‡

August 2025
(working paper)

Abstract

This study investigates how having a comparative advantage in mathematics relative to reading influences educational and career choices, with a particular emphasis on its role in the persistent gender gap in Science, Technology, Engineering, and Mathematics (STEM) fields. Leveraging a rich administrative dataset of students in Catalonia, we track academic performance in math and reading from 6^{th} and 10^{th} grades, along with subsequent enrollment decisions in both high school and university. We find that comparative ability in math accounts for approximately 60% of the gender gap in STEM enrollment at the high school level and 30% at the university level. These findings contribute to the growing literature on gender disparities in STEM by providing novel evidence on the role of relative academic strengths in shaping educational pathways.

Keywords: Gender gap; STEM education; Career choice; Comparative advantage; Academic ability; Educational trajectories; University major choice; Gender disparities; Educational economics.

JEL Codes: I21; I24; J16; J24; D91.

*Institut Català d'Avaluació de Polítiques Públiques (Ivàlua), Institut d'Economia de Barcelona (IEB)-Universitat de Barelona (UB) (email: andreu.arenas@ub.edu)

†Institut d'Anàlisi Econòmica (IAE-CSIC) & Barcelona School of Economics (BSE) (email: lidia.farre@gmail.com)

‡Institut d'Economia de Barcelona (IEB)-Universitat de Barelona (UB) (email: mcramirezmichelena@ub.edu)

Contents

1 Introduction **2**

2 Context and Data **5**

2.1 The Catalan Educational System 5

2.2 Data Availability 6

2.3 Data Descriptives 8

3 Identification Strategy **11**

4 Preliminary Results and Conclusion **14**

4.1 Unconditional on Enrollment Results 15

4.2 Conditional on Enrollment Results 15

1 Introduction

Over the past decades, women have made substantial gains in both educational attainment and labor market participation. Across both developed and developing economies, females now outperform males in overall academic achievement and are more likely to enroll in and complete university degrees. Yet, despite these advances, women remain systematically underrepresented in math-intensive and high-paying fields, particularly within Science, Technology, Engineering, and Mathematics (STEM).

This gender disparity is not uniform across STEM disciplines. While women are the majority in health-related fields, they remain markedly underrepresented in math-intensive areas such as engineering, mathematics, computer science, and the physical sciences. According to recent World Bank (2025), women comprise only 35% of STEM graduates, with particularly low representation in computer science (20%) and engineering (18%).

The underrepresentation of women in STEM has far-reaching consequences for both individuals and society. At the individual level, it contributes to the gender wage gap, as STEM occupations tend to offer significantly higher salaries—on average, two-thirds more—than non-STEM fields (Fry et al. (2021)). These disparities are expected to grow as technological innovation, artificial intelligence, and data analytics increase the demand for STEM-related skills. Consequently, persistent gender segregation in these fields may exacerbate existing income inequalities. Beyond individual outcomes, the scarcity of women in STEM represents a missed opportunity for society: it limits the pool of talent available for innovation, reduces the diversity of perspectives in scientific and technological development, and may constrain aggregate productivity growth.

Despite steady job growth in math-intensive STEM fields, female representation has stagnated. Scholars across disciplines—including biology, psychology, sociology, and economics—have sought to understand the root causes of this gender gap, yet no clear consensus has emerged. Early research attributed the gap to biological differences, arguing that men's superior spatial abilities positioned them to excel in STEM. However, gender disparities in math test scores have narrowed over time, failing to fully explain the persistent underrepresentation of women in STEM. Nonetheless, men continue to outperform women at the highest levels of achievement, which may reinforce the gender gap in STEM participation (Kahn and Ginther (2017); Stoet and Geary (2013)).

Beyond biological explanations, social and psychological factors play a crucial role in shaping women's participation and success in STEM (Kahn and Ginther (2017)). Gender norms and stereotypes, often internalized from a young age, influence beliefs about ability and career preferences, nudging individuals toward traditionally gendered educational paths (Guadalupe (2018); Bottazzi and Lusardi (2020); Breda et al. (2023); Bordalo et al. (2016)). These beliefs

are reinforced by male-dominated cultures, gender discrimination, and the scarcity of female role models in STEM, which can discourage women from pursuing or persisting in math-intensive careers (Bursztyn and Yang (2022)). Psychological dimensions such as self-esteem and confidence further exacerbate these patterns: adolescent boys report higher levels of self-esteem than girls (Bachman et al. (2011)), and math anxiety disproportionately affects girls, undermining their performance and interest in STEM disciplines (Xie et al. (2019)). These early confidence gaps persist into adulthood—women in math-intensive fields tend to underestimate their abilities, while men often overestimate their own and their male peers' performance (Bordalo et al. (2016)).

While psychological and social explanations offer valuable insights, recent contributions in economics have emphasized the role of students' relative abilities in shaping the gender gap in STEM. This literature suggests that academic choices are not solely determined by students' *absolute* proficiency in individual subjects, but are instead strongly influenced by their *comparative advantages* across domains (Wang et al. (2013); Stoet and Geary (2013); Breda and Napp (2019); Goulas et al. (2020)). In the context of STEM, this perspective has received substantial empirical support since girls tend to outperform boys in reading, and boys outperform girls in mathematics. Consequently, even high-achieving female students may rationally opt for non-STEM trajectories, given their relative strength in language-based subjects.

The most influential contribution in this area is Breda and Napp (2019), who analyze data from the 2012 Programme for International Student Assessment (PISA) covering all participating countries. The study examines 15-year-old students' math and reading performance together with their self-reported intentions to pursue math-intensive studies and careers. Intentions are measured through five survey items capturing students' plans to take additional math or science courses, major in math-intensive fields, and pursue math-related occupations. Responses are combined into a single index reflecting the strength of students' math-related intentions. The authors document a significant gender gap in students' intentions to specialize in math, with boys being more likely to choose math-focused fields. Their analysis shows that, across all PISA countries, approximately 78% of the gender gap in intentions to pursue mathematics can be attributed to individual differences in comparative ability between math and reading. This implies that the majority of the gender gap is explained by students' relative advantage in math over reading.

Replicating this analysis for Catalonia using the same PISA 2012 dataset, we find that controlling for comparative advantage in mathematics reduces the gender gap in self-reported intentions to pursue math-related studies by 56%. This reduction is smaller than the cross-country average reported by Breda and Napp (2019), but closely aligned with their estimates for Spain (64%) and France (42%), and lower than those for other Southern European countries such as Italy (95%) and Portugal (89%).

Building on these insights, our study investigates the role of absolute ability and comparative advantage in shaping students' educational and career choices. Using administrative data from Catalonia (one of the largest regions in Spain), we examine how proficiency in math and reading contributes to explaining the gender gap in math-intensive fields. Specifically, we use standardized test scores from the end of primary (6^{th} grade) and lower secondary education (10^{th} grade) to analyze how early academic strengths relate to students' subsequent track choices in upper secondary school (11^{th} and 12^{th} grades) and their selection of university majors.

Our findings are broadly consistent with those of the literature, reinforcing the role of comparative advantage in shaping gendered educational choices. Specifically, our analysis suggests that differences in comparative ability explain approximately 60% of the gender gap in STEM enrollment at the high school level and 30% at the university level. The discrepancy in explanatory power between the high school and university levels may reflect both the limitations of using self-reported intentions at age 15 as a proxy for actual enrollment decisions, and the possibility that comparative advantage plays a more limited role in later educational choices than initially assumed.

We contribute to the literature by providing new evidence on the role of absolute and comparative advantage in shaping educational and career choices, particularly in the decision between math-intensive and non-math-intensive pathways. Our analysis traces the impact of these abilities across different stages of schooling, drawing on standardized test data from both primary and lower secondary education. We extend previous work by exploring heterogeneous effects across gender and ability levels, and by examining variation in the math intensity of university majors. This allows for a more nuanced understanding of enrollment into STEM-intensive fields—those characterized by both high labor market returns and persistent female underrepresentation. To capture this variation, we report results using multiple definitions of STEM. Furthermore, we address a key limitation in Breda and Napp (2019) by relying on administrative records of actual enrollment decisions rather than self-reported intentions.

In ongoing work, we extend the analysis to incorporate the role of *relative* academic ability—measured with respect to school peers—in shaping students' educational trajectories. This addition enables us to assess whether a student's rank within the classroom distribution influences track and field choices, beyond the effects of absolute and comparative skills.

Motivated by the promising role of comparative advantage in career decision-making, we will implement a low-cost, scalable information intervention designed to promote gender equity in STEM. The intervention, structured as a randomized controlled trial (RCT), is scheduled for the beginning of the 2025–2026 academic year (September/October 2025) and will reach approximately 2,500 middle school students in Catalonia. By raising students' awareness of the

persistent gender gap in STEM and providing information on the skills required to succeed in these fields, the intervention aims to empower students to make informed academic decisions aligned with their strengths and interests, ultimately contributing to greater female representation in math-intensive careers.

Ongoing analyses, including those on relative ability within classrooms and from the randomized controlled trial, will be integrated into a future version of this paper.

2 Context and Data

2.1 The Catalan Educational System

Catalonia provides a relevant setting for studying educational and occupational choices in developed economies. As one of Spain's most affluent and industrialized regions, it features a diversified economy, high levels of educational attainment, and a labor market increasingly oriented toward knowledge-intensive sectors. These characteristics make Catalonia broadly comparable to other high-income European regions, supporting the external validity of our findings beyond the Spanish context.

As in most Western countries, compulsory education in Catalonia comprises primary and lower secondary schooling, spanning ten years in total (from ages 6 to 16). Before completing 10^{th} grade, students who wish to continue with their high school studies and pursue upper secondary education (i.e., grades 11 and 12, jointly known in Catalonia as *Batxillerat*) must choose a high school specialization track that will apply for both years. The available tracks are (i) Arts, (ii) Humanities and Social Sciences, and (iii) Sciences and Technology, with the latter encompassing both STEM-oriented and health-related fields.[1].

Admission to STEM programs at Spanish universities typically requires prior enrollment in the Sciences and Technology track. Students who did not follow this track must take additional subject-specific entrance exams in mathematics and science to demonstrate sufficient proficiency. This alternative pathway is considerably more demanding, as it requires students to prepare independently for exams without having completed the corresponding upper secondary coursework. Unsurprisingly, very few students pursue this route: in our data, only 1.3% of university entrants who enrolled in STEM programs had not previously followed the Sciences and Technology track. Although we do not directly observe whether students took the required exams, we infer this pathway indirectly through mismatches between upper secondary track and university major. This institutional structure implies that early specialization plays a decisive role in shaping students' academic and career trajectories. Consequently, access to STEM fields is largely determined well before university entry.

[1]Since 2022, a General high school track has been introduced in some Catalan schools. This more flexible pathway combines elements from different specializations, but it is only available in a limited number of schools and affects only the youngest cohort in our sample.

2.2 Data Availability

We use detailed administrative records from the Statistical Institute of Catalonia (IDESCAT), covering a representative sample of 10^{th} grade students in Catalonia (i.e., those in their final year of compulsory education) born between 2000 and 2007 (219,637 individuals in total). The dataset provides rich, student-level panel information, including: (i) socioeconomic background, (ii) standardized test scores from both 6^{th} and 10^{th} grades, and (iii) educational enrollment decisions following completion of compulsory education (if any). Each of these components is detailed below.

First, we observe students' basic socio-demographic characteristics at the end of compulsory education. The data include each student's age, gender, and country of origin, as well as the education level, country of origin, and gender of their tutors. In addition, we record an anonymized school identifier and the school's public or private status.

Secondly, our dataset contains 6^{th} and 10^{th} grade math and reading[2] standardized test scores. These tests are designed by the Catalan educational authorities and administered on the same date across all schools, public and private alike. They are mandatory and contribute to students' final GPA.

Following standard practice in the literature (Breda and Napp, 2019), we use math and reading scores to construct a measure of students' comparative advantage in mathematics, defined as the individual difference between their standardized math and reading scores. A positive value indicates that the student performs better in math relative to reading. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one (see Figure 1 for the distribution of these scores by gender).

Lastly, we observe students' educational trajectories following the completion of compulsory education, including whether they enrolled in upper secondary education (grades 11 and 12) and whether they later pursued university studies. Based on these enrollment choices, we construct a series of STEM indicators that capture academic specialization at both the upper secondary and university levels. We define a **high school STEM indicator** which equals one if the student enrolled in the Sciences and Technology track during grades 11 and 12. While this track provides the standard preparation for STEM degrees, it also serves as the primary entry route into life sciences and health-related programs, which are not typically considered as STEM in the literature. As such, this indicator reflects exposure to a math- and science-intensive curriculum, and should not be interpreted as a measure of commitment to math-intensive fields alone.

[2]Catalonia is a bilingual region where both Spanish and Catalan are part of the school curriculum, and students are required to take standardized tests in each language separately. We define reading performance as the higher of the two scores obtained in these tests. As a robustness check, we replicate our analysis using the average score across Spanish and Catalan, as well as using each language individually. Findings remain consistent across all specifications.

At the university level, we construct three distinct measures of STEM enrollment. The first is a **standard university STEM indicator**, which captures students enrolled in majors traditionally classified as STEM in the academic literature. This category encompasses degrees in mathematics, computer science, engineering, and information and communication technologies (ICT). This definition excludes programs in medicine and the life sciences, as these fields are less male-dominated and are therefore typically not included under the STEM umbrella in studies examining gender disparities. The second is a **broad university STEM indicator**, which further extends STEM definition to include life sciences, health-related disciplines, and agriculture; aligning more closely with the scope of the Sciences and Technology high school track curricula (i.e., our high school STEM indicator). The third is a more **restrictive university STEM indicator**, which focuses exclusively on the most quantitatively rigorous and math-intensive disciplines, focusing on fields that heavily rely on mathematics, physics, and formal analytical skills. These three definitions allow us to capture meaningful variation in the composition of STEM fields in terms of mathematical intensity, expected earnings, and gender representation. A detailed mapping of university majors included in each category is provided in Appendix Table A.1. Throughout the analysis, we report results using all three university-level STEM definitions to ensure robustness and to account for different forms of specialization.

As compulsory education in Catalonia ends at 10^{th} grade, not all students in our sample continue with upper secondary or university studies. Consequently, the sample composition evolves across educational stages: starting from the full sample of 10^{th} grade students, narrowing to those who enroll in upper secondary education (approximately 70% of students completing compulsory education), and further to those who eventually enroll in university (around 45% students completing compulsory education). Moreover, information on university enrollment is only available for the older cohorts (born between 2000 and 2003), as younger cohorts have either not yet reached university age or administrative data is not yet available. Table 1 provides an overview of data availability by student birth cohort.

Given this structure of sample composition across educational stages, STEM indicators are defined consistently but their interpretation varies by sample. In the full sample of 10^{th}grade students (i.e., the *unconditional sample)*, a value of one indicates enrollment in a STEM track or major, while zero reflects either a non-STEM choice or leaving the education system. In the *sample conditional on upper secondary enrollment*, the high school indicator equals one for students in the Sciences and Technology track and zero otherwise; university-level indicators take the value zero both for students in non-STEM majors and for those not continuing to university. In the *sample conditional on university enrollment*, all students have advanced through both stages, so STEM indicators equal zero only for non-STEM majors. Analyses of university outcomes are further restricted to cohorts born between 2000 and 2003, as data for the younger cohorts is not available.

Table 1: Sample Composition by Student Birth-Cohort

	Compulsory Education		**Post-Compulsory Education**	
Students' birth year	6^{th} grade test scores	10^{th} grade test scores	High school enrollment track	University enrollment major
2000	✓	✓	✓	✓
2001	✓	✓	✓	✓
2002	✓	✓	✓	✓
2003	✓	✓	✓	✓
2004	✓	✓	✓	✗
2005	✓	✓	✓	✗
2006	✓	✓	✓	✗
2007	✓	✓	✓	✗
Obs.	219637		155494	45606

Note: A checkmark indicates the availability of the corresponding information for each birth cohort. 6^{th} and 10^{th} grade test scores and upper secondary enrollment choices are available for all cohorts, whereas university enrollment choices are observed only for the older cohorts. Source: IDESCAT data.

2.3 Data Descriptives

Table 2 presents descriptive statistics for the full sample of students completing compulsory education, additional statistics are detailed for subsamples conditional on enrollment in upper secondary (Table A.2) and university (Table A.3). As we move from the full sample of students completing compulsory education to those pursuing post-compulsory studies, several systematic patterns emerge: the proportion of female students increases, the share of foreign-born students declines, parental educational attainment rises, and enrollment in STEM tracks becomes more prevalent[3].

Gender differences in measured abilities and student characteristics are reported in Table 3. Tables A.6 and A.7 present analogous statistics for the subsamples restricted to students enrolled in upper secondary education and university, respectively. Two primary gender-related dimensions of variation emerge: differences in academic performance across domains, and differences in educational enrollment patterns.

Focusing first on performance differences, prior evidence indicates that aggregate gender gaps in math and reading scores have narrowed in most countries, including Catalonia. However, while gender gaps in math ability remain modest around the middle of the distribution, they widen substantially at the upper tail, where boys outperform girls at the highest levels of achievement. In contrast, the opposite pattern emerges in reading, where girls consistently outperform boys.

Our analysis of standardized test scores in 6^{th} and 10^{th} grade reveals two important dynamics

[3]See in the Appendix Tables A.4 and A.5 and Figure A.5 for key differences in student characteristics and changes in performance distribution across the unconditional sample and samples conditional on upper secondary and university enrollment, respectively.

in math and reading achievement. First, gender differences are asymmetric: the male advantage in math is larger than the female advantage in reading. Second, this disparity widens over time, increasing from 0.36 standard deviations in 6^{th} grade to 0.59 in 10^{th} grade. The corresponding histogram and kernel distributions are presented in Appendix Figures A.1 and A.2.

These dynamics also evolve across the ability distribution. In 6^{th} grade, boys and girls perform similarly in reading conditional on math ability, while boys hold only a modest edge in math. By 10^{th} grade, however, male students consistently outperform female students in math across the entire distribution, with the gap widening at the upper tail. Girls retain an advantage in reading, but it remains smaller in magnitude than the male advantage in math. Appendix Figures A.3 and A.4 illustrate these patterns, which underscore a growing divergence in comparative advantages during adolescence, likely shaping subsequent educational choices.

Raw test scores (0–100 scale) reveal an additional dimension: reading scores are, on average, higher than math scores for both genders and across both grades. In 6^{th} grade, the average score in both subjects is around 80 points, with gender gaps of approximately 3 points in math (favoring boys) and 2 points in reading (favoring girls). By grade 10, average math scores decline to 68 points while reading scores remain high at 81. At this stage, boys outperform girls in math by 4.8 points, while girls outperform boys in reading by 3.7 points. In absolute terms, the male advantage in math is larger than the female advantage in reading. However, because reading scores have a higher mean and lower variance, standardized scores provide a more comparable measure of relative performance across domains.

Incorporating math comparative advantage provides additional insights into students' skill profiles. As shown in Figure 1, comparative advantage in math is closely tied to math ability rather than to reading skills, and this pattern is consistent across genders. Yet among high-achieving math students, girls display comparatively stronger reading abilities than boys with similar levels of math performance.

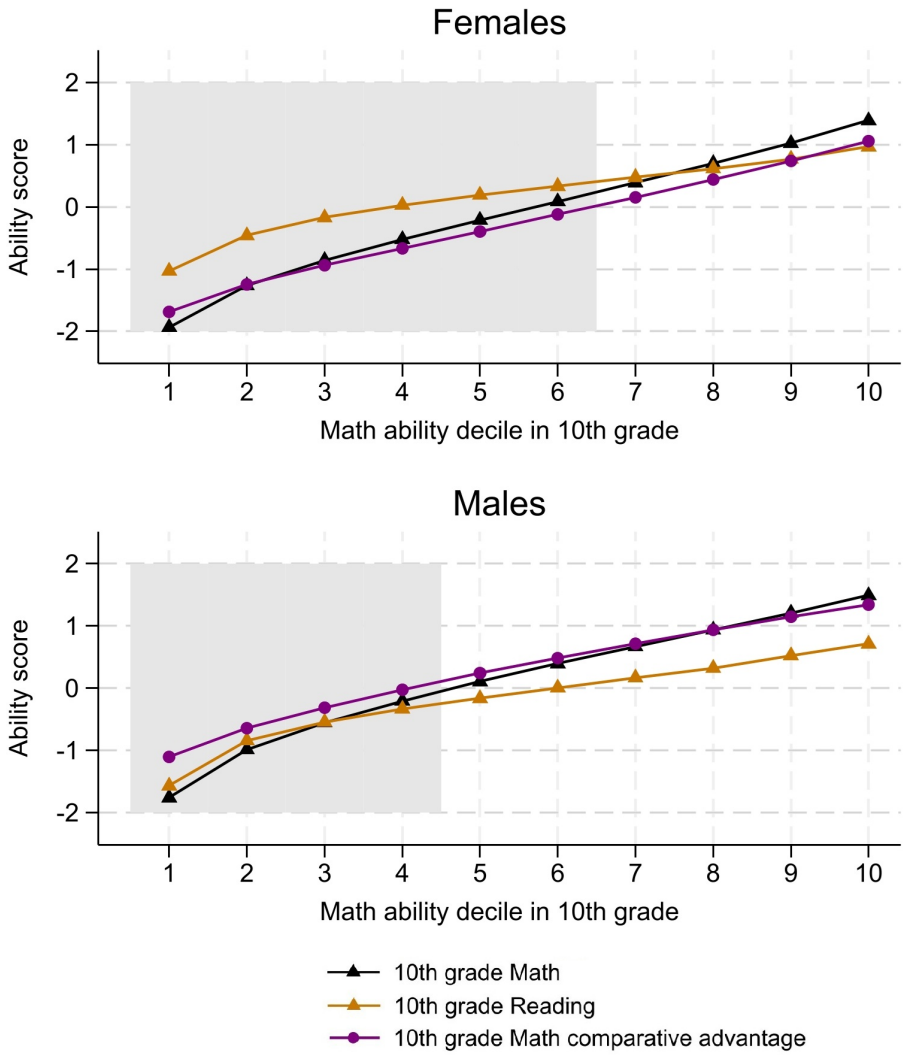


Figure 1: **10^{th}-Grade Abilities, by Gender.** The figure shows the distribution of standardized test scores in math, reading, and math comparative advantage separately for female students (top panel) and male students (bottom panel). All variables are standardized to have mean zero and standard deviation one. The gray background indicates math ability deciles in which students have a reading advantage compared to math. The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment. Source: Own elaboration based on IDESCAT data.

Building on these ability differences, we now examine how they map into educational choices. Figure 2 documents the gender disparities in STEM participation across high school and university. The Y-axis shows the proportion of female students enrolling in STEM tracks or majors relative to all female students at each stage, while the X-axis shows the corresponding proportions for male students. Equal participation would place all points along the 45-degree diagonal; instead, all points fall below this line, indicating consistently lower female representation. In high school, 40% of female students opt for the Science and Technology track, compared to 60% of male students (roughly two females for every three males). At the university level, the gap widens as the definition of STEM becomes narrower. Including life sciences and health-related majors, 30% of female students enroll in STEM compared to 50% of males. Excluding these

fields, female representation drops to just over 10%, versus 40% for males, which corresponds to a ratio exceeding three to one. Shares are comparable under both the standard STEM classification and the math-intensive classification.

These patterns highlight that gender gaps in STEM enrollment cannot be understood solely as a matter of choice; they are intertwined with differences in academic performance profiles, particularly in math-intensive domains, which may influence both perceived and actual preparedness for STEM pathways.

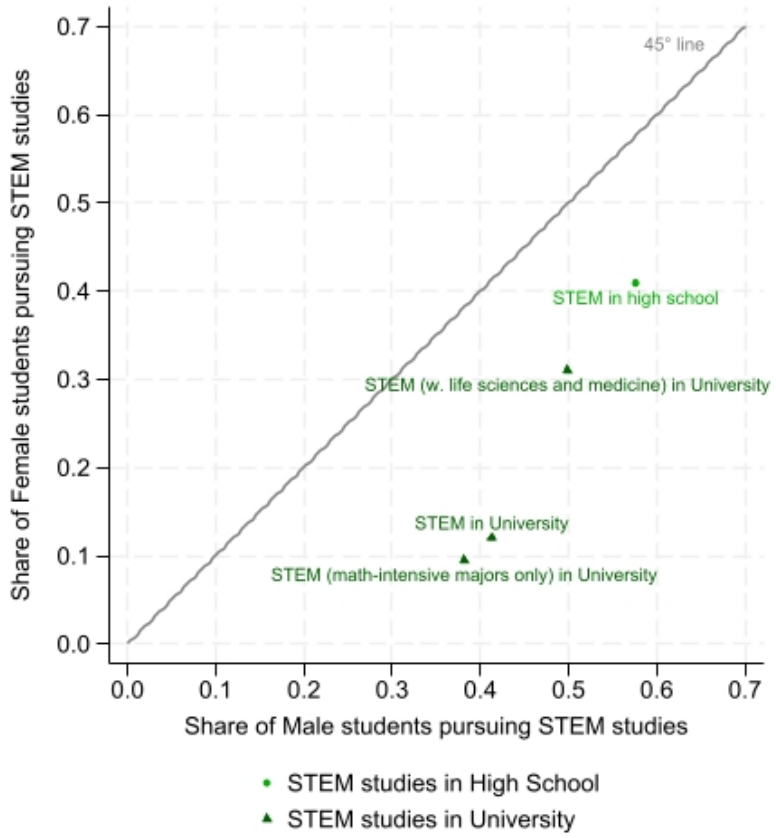


Figure 2: **Share of Male and Female Students Pursuing STEM Tracks in High School and University.** Each axis shows, for each gender, the share of students enrolling in a STEM track in high school and in a STEM major at university, each calculated relative to the total number of enrollees at the corresponding level. This disparity further increases in math-intensive majors. While male and female students participate in STEM tracks at closer rates in high school, the transition to university exacerbates gender gaps, especially in math-intensive fields. This reflects both the overall underrepresentation of women in STEM and their uneven distribution across STEM subfields. Source: Own elaboration based on IDESCAT data.

3 Identification Strategy

Our baseline empirical model is estimated using the following linear specification:

$$STEM_{isc} = \alpha \cdot Female_i + \beta \cdot Ability_i + \delta' \cdot X_i + \lambda_s + \gamma_c + \epsilon_{isc}$$

The dependent variable, $STEM_{isc}$, is a binary indicator equal to one if student i, from school s and cohort c, enrolls in a STEM track, either in high school or university. The key independent variable, $Female_i$, is an indicator for gender, that equals one for female students. $Ability_i$ includes either math scores, reading scores, or the difference between the two (as a measure of comparative advantage in math), measured both in grades 6^{th} and 10^{th}. We estimate separate models for each ability measure and grade to examine how academic strengths mediate gender differences in STEM enrollment. We do not include all three ability measures simultaneously, as the comparative advantage variable is mechanically defined as the difference between math and reading scores. Including all three would result in perfect multicollinearity and complicate the interpretation of coefficients. To avoid this, we estimate separate specifications for each measure - math, reading, and math comparative advantage - highlighting their individual explanatory power.

The vector X_i includes a set of student background characteristics: whether the student is foreign-born, lives in a two-tutor household, has same-gender tutors, whether at least one tutor holds a college degree, and whether one or both tutors are foreign-born. We also include fixed effects for school (λ_s) and cohort (γ_c) to control for time-invariant unobserved heterogeneity at these levels, and report heteroskedasticity-robust standard errors.

This empirical framework quantifies the role of academic ability and student background characteristics in shaping STEM enrollment decisions, while flexibly accounting for institutional and temporal heterogeneity. By distinguishing between high school and university outcomes, as well as between abilities measured at the end of primary and middle school, we provide evidence on to which extent early skill profiles influence long-term educational trajectories.

We estimate the model for three samples: (i) the unconditional sample, which includes all students who completed compulsory education from cohorts 2000 to 2007 (regardless of whether they pursued further education after completing 10^{th} grade); (ii) a sample conditional on enrollment in upper-secondary education; and (iii) a sample conditional on university enrollment.

The first unconditional approach mirrors the analysis conducted by Breda and Napp (2019) using PISA 2012 data: we exploit abilities in 10^{th} (where students are on average 15 years old) to explain their subsequent study choices. The key difference between the two approaches is that we are able to rely on actual enrollment decisions rather than self-reported intentions. It is important to note, however, that exploiting the unconditional sample does not allow us to isolate the effect of having a math comparative advantage on the decision to pursue STEM studies from the broader decision of continuing their education. As such, this specification does not strictly examine the gender gap in STEM enrollment among students who remain in the system, but instead captures variation in both educational continuation and field choice.

To address this limitation, we also estimate models conditioning on enrollment. This approach shifts the focus from predicting whether students continue beyond compulsory education to, conditional on continuation, examining the gender gap in STEM-oriented academic decisions. Specifically, we analyze gender differences in track selection among upper-secondary students and major selection among university students.

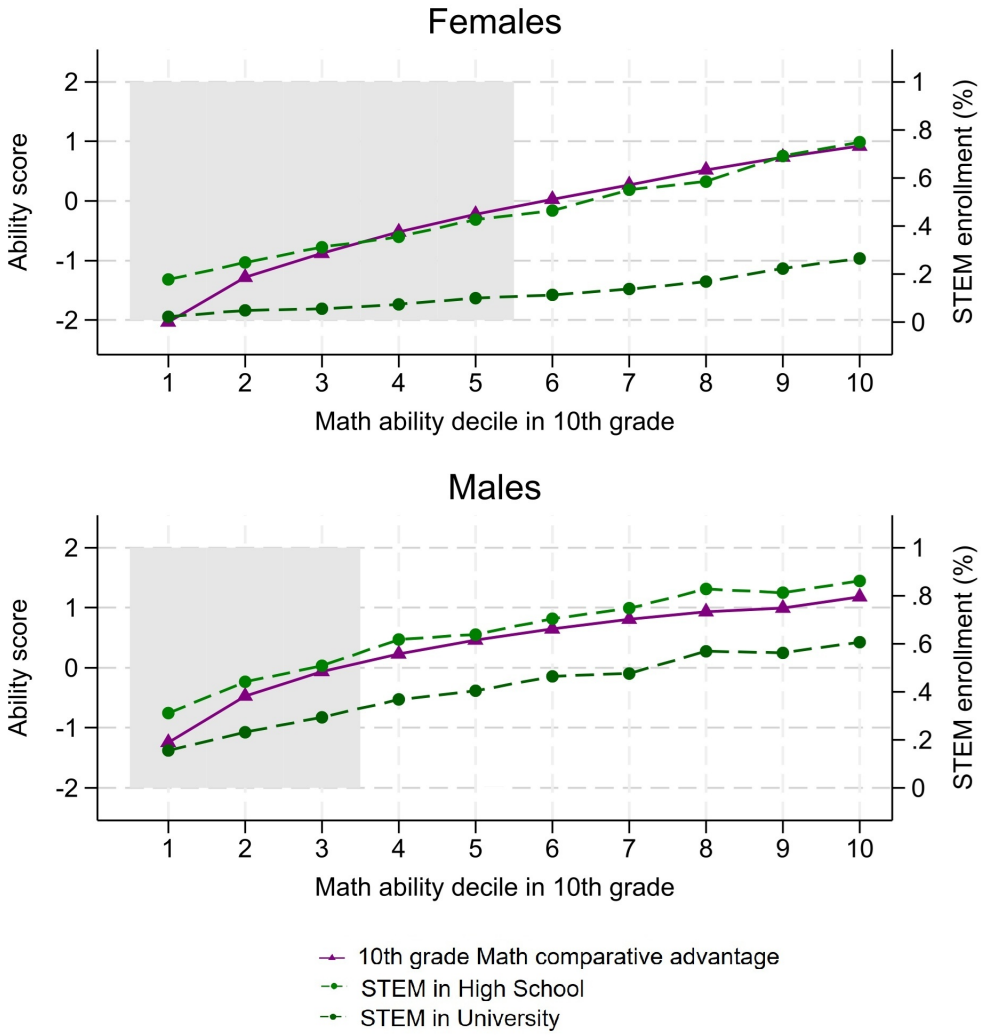


Figure 3: **STEM choices and Math Comparative Advantage by 10^{th}-grade Math Ability Decile, by Gender.** The figure shows, for the sample of students conditional on university enrollment, the average math comparative advantage (left axis) together with the share of students enrolled in STEM tracks at the upper secondary and university levels (right axis). Results are presented separately for female students (top panel) and male students (bottom panel). All measures are plotted across deciles of 10^{th}-grade math performance. The gray background indicates math ability deciles in which students have a reading advantage compared to math. Source: Own calculations based on IDESCAT data.

Lastly, Figure 3 illustrates the intuition behind the role of math comparative advantage in shaping study choices. For students who continued to upper secondary and university, we plot their math comparative advantage scores together with the share who pursue STEM tracks in high school and university, separately by gender. A clear pattern emerges: male students display

a positive comparative advantage in math already from the third decile of performance onward, whereas for female students this only becomes evident at much higher levels of the distribution. Consistent with this, female students enroll in the high school STEM track at substantially lower rates across almost the entire ability spectrum, while male students appear to over-select into STEM irrespective of their actual performance.

Turning to university enrollment, the gender gap widens even further, and the role of math comparative advantage becomes less apparent. Among female students, STEM participation remains low across the ability spectrum, with only modest increases at the very top deciles. By contrast, male students exhibit consistently higher STEM enrollment rates at all performance levels, and their propensity to choose STEM appears only weakly tied to their comparative advantage in math. This pattern reinforces the idea that women face substantial barriers to entering STEM even when they hold a strong relative advantage, while men tend to self-select into STEM fields broadly.

4 Preliminary Results and Conclusion

This section presents preliminary evidence on how absolute and comparative academic abilities influence students' decisions to pursue STEM tracks at both the high school and university levels, using both unconditional (Tables 4 to 7) and conditional (Tables 8 to 15) samples.

Across all specifications, we find that academic skills measured in 6^{th} grade are less predictive of STEM enrollment than those assessed at the end of compulsory education (10^{th} grade). This highlights middle school as a pivotal period in shaping students' academic trajectories.

Nonetheless, there is a strong positive correlation in academic performance between 6^{th} and 10^{th} grades for both girls and boys, indicating substantial stability in students' relative academic standing over time. Figures A.6 and A.7 illustrate this persistence in math and reading performance, respectively, disaggregated by gender. Notably, while performance trajectories are nearly identical for males and females in reading, small gender differences emerge in math. In particular, female students in the top decile of math ability in 6^{th} grade tend to slightly outperform their male peers by 10^{th} grade. Further analysis will be conducted to account for this persistence.

Together, these findings suggest a dynamic developmental process: relative strengths that emerge during primary school—especially in math and reading—can lay the foundation for later educational choices. For simplicity and comparability, the remainder of the analysis focuses on academic abilities as measured at the end of middle school.

4.1 Unconditional on Enrollment Results

The raw gender gap in STEM enrollment is approximately 5% at the high school level and widens to about 10% at the university level for STEM and math-intensive university careers (see Tables 4 to 7, which report estimates for all STEM enrollment across high school and university levels, using the unconditional sample of students completing compulsory education).

As shown in Table 4, controlling for students' math comparative advantage in 10^{th} grade markedly reduces the gender gap in high school STEM enrollment, flipping the female coefficient from -0.05 to $+0.03$ percentage points. At the university level, Table 5 shows that conditioning on comparative advantage lowers the gender gap in STEM by nearly forty percent, from -0.098 to -0.064 percentage points.

These results underscore the explanatory power of math comparative advantage relative to absolute math and reading abilities. Yet, an important difference with respect to the literature emerges: the unconditional gender gap in our context is relatively modest (about 5–10%), substantially smaller than the gap in intentions to pursue math-intensive studies reported by Breda and Napp (2019) for Spain (around 80%). This discrepancy suggests that self-reported intentions in PISA tests may not fully translate into actual educational choices.

Taken together, these findings indicate that comparative advantage in mathematics not only predicts STEM track selection, but also shapes broader educational trajectories at both the upper-secondary and tertiary levels.

4.2 Conditional on Enrollment Results

Conditioning on students enrolled in a high school track, the raw gender gap in STEM choice is about 16% at both the high school and university levels. Controlling for students' math comparative advantage substantially reduces the gap in high school by nearly 60%, from 0.16 to 0.06 percentage points (see Table 8). At the university level, however, the reduction is more modest, with the gap narrowing by about 30%, from 0.16 to 0.11 percentage points for standard STEM majors (see Table 9).

The gender disparity is most pronounced when conditioning on university enrollment, where the raw gap reaches 19% in high school STEM choice and nearly 30% for standard STEM majors at the tertiary level. Adjusting for math comparative advantage reduces the high school gap by about 50% and the university gap by roughly 30% (see Tables 12 and 13).

While the literature emphasizes that absolute ability cannot account for gender disparities in STEM as effectively as math comparative advantage, our findings indicate that absolute math ability still exerts a meaningful influence on STEM enrollment. We cannot rule out that its effect differs significantly from that of comparative advantage, particularly when conditioning

the sample on enrollment. To further investigate, we employ an Oaxaca-Blinder decomposition to disentangle the contributions of absolute and comparative math abilities to STEM entry decisions. Both dimensions of math ability are associated with a higher probability of entering STEM for male students than for female students. This suggests that female students with similar skill profiles are less likely to translate those abilities into STEM choices. Although absolute math ability yields a larger marginal return overall, gender disparities are sharper in math comparative advantage, making it the primary driver of the observed gender gap in STEM entry.

When examining heterogeneity by the breadth of the university STEM definition, we find that both the raw gender gap and the contribution of math comparative advantage are nearly identical under the standard and the more restrictive math-intensive definition, regardless of whether we condition on high school or university enrollment (see Tables 11 and 15). By contrast, when life sciences and medicine are included, the raw gender gap diminishes substantially, and so does the explanatory power of abilities.

Moreover, in Tables 16 and 17, we replicate the analysis for the sample conditional on university enrollment, this time stratifying by students' 10^{th}-grade math performance. The results show that the effect of math comparative advantage is concentrated among students in the second-highest quartile of math ability. This suggests that the students most responsive to their relative strengths are not necessarily the top performers - who systematically choose STEM - but rather those just below the top quartile. These are students with above-average math abilities who could have enrolled in STEM studies and pursued a potentially successful career, but instead opted for other paths. This finding highlights a critical policy implication: interventions designed to reduce the gender gap in STEM may be most effective when targeted at this group of capable but undecided students, rather than exclusively at top performers.

These findings carry important implications for the design of interventions aimed at reducing the gender gap in STEM enrollment in a cost-effective and scalable way. While broader efforts to reshape gender norms and career expectations remain essential, such approaches are typically costly and slow to generate measurable impact. By contrast, a low-cost and targeted intervention that emphasizes the role of comparative advantage could offer a practical strategy for narrowing gender disparities in STEM participation - by as much as 50-to-60% in high school and around 30% in university STEM careers, including those that are more math-intensive.

In sum, our results confirm the importance of math comparative advantage as a mechanism shaping educational choices related to STEM. While absolute math ability remains a significant predictor — particularly at the university level — math comparative advantage offers a more nuanced explanation of gender gaps. Future work will explore these dynamics in greater depth and evaluate the potential of targeted interventions to promote more equitable academic trajectories.

Tables

Table 2: Descriptive statistics
Unconditional sample

Variable	Mean	SD	Min	Max	Obs.
Female	0.51	0.50	0.00	1.00	219637
6th grade Math	0.00	1.00	-5.33	1.70	219637
6th grade Reading	0.00	1.00	-6.19	1.67	219637
6th grade math comparative advantage	0.00	1.00	-6.39	5.13	219637
10th grade Math	0.00	1.00	-3.94	1.85	219637
10th grade Reading	0.00	1.00	-8.36	1.86	219637
10th grade math comparative advantage	0.00	1.00	-5.45	6.49	219637
Birth year	2003.58	2.27	1998.00	2008.00	219637
Foreign student	0.07	0.25	0.00	1.00	219637
Two-member household	0.90	0.30	0.00	1.00	219637
Same-sex household	0.01	0.11	0.00	1.00	219637
HH with both tutors college-educated	0.19	0.39	0.00	1.00	219637
HH with only one tutor college-educated	0.22	0.41	0.00	1.00	219637
HH with only one tutor foreign-born	0.13	0.34	0.00	1.00	219637
HH with both tutors foreign-born	0.11	0.31	0.00	1.00	219637
STEM in high school	0.34	0.47	0.00	1.00	219637
STEM in university	0.11	0.31	0.00	1.00	103951
STEM (w. life sciences and medicine) in university	0.17	0.38	0.00	1.00	103951
STEM (math-intensive majors only) in university	0.10	0.30	0.00	1.00	103951

Note: The sample is representative of 10th grade students in Catalonia and is unconditional on post-compulsory education enrollment. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are only observed for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table 3: Descriptive Statistics by Gender
Unconditional Sample

Variable	Diff./SE	Females	Males	p-value
6th grade Math	0.20***	-0.10	0.10	0.00
	(0.00)			
6th grade Reading	-0.16***	0.08	-0.08	0.00
	(0.00)			
6th grade Math comparative advantage	0.40***	-0.20	0.20	0.00
	(0.00)			
10th grade Math	0.25***	-0.12	0.13	0.00
	(0.00)			
10th grade Reading	-0.35***	0.17	-0.18	0.00
	(0.00)			
10th grade Math comparative advantage	0.54***	-0.27	0.27	0.00
	(0.00)			
Birth year	0.03***	2003.57	2003.60	0.00
	(0.01)			
Foreign student	-0.01***	0.07	0.06	0.00
	(0.00)			
Two-member household	0.01***	0.90	0.91	0.00
	(0.00)			
Same-sex household	0.00	0.01	0.01	0.19
	(0.00)			
HH with both tutors college-educated	0.01***	0.19	0.20	0.00
	(0.00)			
HH with only one tutor college-educated	0.01***	0.21	0.22	0.00
	(0.00)			
HH with only one tutor foreign-born	-0.01***	0.13	0.13	0.00
	(0.00)			
HH with both tutors foreign-born	-0.01***	0.11	0.11	0.00
	(0.00)			
STEM in high school	0.05***	0.32	0.37	0.00
	(0.00)			
Observations	219637			
STEM in university	0.10***	0.06	0.16	0.00
	(0.00)			
STEM (w. life sciences and medicine) in university	0.04***	0.16	0.19	0.00
	(0.00)			
STEM (math-intensive majors only) in university	0.10***	0.05	0.15	0.00
	(0.00)			
Observations	103951			

Note: The sample is representative of 10th grade students in Catalonia and is unconditional on post-compulsory education enrollment. Column "Diff./SE" reports the difference in means between females and males, with standard errors in parentheses. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are available exclusively for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table 4: Estimation Results for pursuing a STEM Track in High School
Unconditional sample

	(1) STEM in High School	(2) STEM in High School	(3) STEM in High School	(4) STEM in High School	(5) STEM in High School	(6) STEM in High School	(7) STEM in High School	(8) STEM in High School
Female	-0.0548***	-0.0520***	-0.0241***	-0.0690***	-0.0262***	-0.00370*	-0.102***	0.0306***
	(0.00202)	(0.00198)	(0.00190)	(0.00194)	(0.00200)	(0.00182)	(0.00193)	(0.00195)
6th grade Math			0.143***					
			(0.000880)					
6th grade Reading				0.101***				
				(0.000936)				
6th grade Math comparative advantage					0.0641***			
					(0.000928)			
10th grade Math						0.205***		
						(0.000891)		
10th grade Reading							0.138***	
							(0.000925)	
10th grade Math comparative advantage								0.154***
								(0.000951)
Constant	0.370***	0.252***	0.276***	0.286***	0.243***	0.298***	0.317***	0.245***
	(0.00147)	(0.00415)	(0.00400)	(0.00410)	(0.00411)	(0.00383)	(0.00406)	(0.00396)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	219637	219635	219635	219635	219635	219635	219635	219635
r2	0.003	0.068	0.149	0.109	0.085	0.223	0.139	0.158

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing the STEM track in high school. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 5: Estimation Results for pursuing a STEM Major in University
Unconditional sample

	(1) STEM in University	(2) STEM in University	(3) STEM in University	(4) STEM in University	(5) STEM in University	(6) STEM in University	(7) STEM in University	(8) STEM in University
Female	-0.0987***	-0.0970***	-0.0855***	-0.105***	-0.0874***	-0.0792***	-0.117***	-0.0645***
	(0.00192)	(0.00190)	(0.00184)	(0.00191)	(0.00192)	(0.00180)	(0.00196)	(0.00186)
6th grade Math			0.0597***					
			(0.000836)					
6th grade Reading				0.0500***				
				(0.000857)				
6th grade Math comparative advantage					0.0239***			
					(0.000787)			
10th grade Math						0.0872***		
						(0.000972)		
10th grade Reading							0.0580***	
							(0.000909)	
10th grade Math comparative advantage								0.0672***
								(0.000919)
Constant	0.159***	0.111***	0.119***	0.128***	0.106***	0.128***	0.137***	0.107***
	(0.00162)	(0.00368)	(0.00364)	(0.00369)	(0.00367)	(0.00360)	(0.00372)	(0.00361)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	103951	103947	103947	103947	103947	103947	103947	103947
r2	0.025	0.072	0.105	0.094	0.078	0.138	0.102	0.112

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: pursuing a standard STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 6: Estimation Results for Pursuing a Broadly-Defined STEM Major in University
Unconditional Sample

	(1) STEM (Broad) in University	(2) STEM (Broad) in University	(3) STEM (Broad) in University	(4) STEM (Broad) in University	(5) STEM (Broad) in University	(6) STEM (Broad) in University	(7) STEM (Broad) in University	(8) STEM (Broad) in University
Female	-0.0358***	-0.0331***	-0.0159***	-0.0458***	-0.0199***	-0.00624**	-0.0642***	0.0150***
	(0.00236)	(0.00230)	(0.00224)	(0.00227)	(0.00234)	(0.00217)	(0.00230)	(0.00229)
6th grade Math			0.0898***					
			(0.00102)					
6th grade Reading				0.0786***				
				(0.00104)				
6th grade Math comparative advantage					0.0331***			
					(0.000974)			
10th grade Math						0.132***		
						(0.00115)		
10th grade Reading							0.0912***	
							(0.00110)	
10th grade Math comparative advantage								0.0995***
								(0.00111)
Constant	0.193***	0.116***	0.128***	0.144***	0.110***	0.142***	0.157***	0.111***
	(0.00175)	(0.00434)	(0.00425)	(0.00432)	(0.00433)	(0.00418)	(0.00434)	(0.00423)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	103951	103947	103947	103947	103947	103947	103947	103947
r2	0.002	0.077	0.126	0.113	0.084	0.178	0.125	0.136

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a broadly-defined (including life sciences and medicine) STEM major in university. Controls: Student's birth year, foreign student, status two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 7: Estimation Results for Pursuing a Restrictively-Defined STEM Major in University
Unconditional Sample

	(1) STEM (Rest.) in University	(2) STEM (Rest.) in University	(3) STEM (Rest.) in University	(4) STEM (Rest.) in University	(5) STEM (Rest.) in University	(6) STEM (Rest.) in University	(7) STEM (Rest.) in University	(8) STEM (Rest.) in University
Female	-0.0996***	-0.0981***	-0.0879***	-0.105***	-0.0897***	-0.0821***	-0.116***	-0.0691***
	(0.00182)	(0.00180)	(0.00175)	(0.00182)	(0.00182)	(0.00171)	(0.00188)	(0.00175)
6th grade Math			0.0536***					
			(0.000794)					
6th grade Reading				0.0455***				
				(0.000816)				
6th grade Math comparative advantage					0.0210***			
					(0.000743)			
10th grade Math						0.0787***		
						(0.000931)		
10th grade Reading							0.0531***	
							(0.000870)	
10th grade Math comparative advantage								0.0601***
								(0.000872)
Constant	0.147***	0.105***	0.112***	0.121***	0.101***	0.121***	0.129***	0.102***
	(0.00157)	(0.00351)	(0.00347)	(0.00353)	(0.00350)	(0.00345)	(0.00355)	(0.00344)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	103951	103947	103947	103947	103947	103947	103947	103947
r2	0.028	0.071	0.100	0.091	0.076	0.131	0.098	0.107

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a restrictively-defined (math-intensive majors only) STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 8: Estimation Results for pursuing a STEM Track in High School
Conditional on Upper Secondary Enrollment Sample

	(1) STEM in High School	(2) STEM in High School	(3) STEM in High School	(4) STEM in High School	(5) STEM in High School	(6) STEM in High School	(7) STEM in High School	(8) STEM in High School
Female	-0.167***	-0.162***	-0.127***	-0.168***	-0.130***	-0.0933***	-0.187***	-0.0648***
	(0.00251)	(0.00253)	(0.00250)	(0.00251)	(0.00257)	(0.00244)	(0.00250)	(0.00255)
6th grade Math			0.121***					
			(0.00120)					
6th grade Reading				0.0603***				
				(0.00127)				
6th grade Math comparative advantage					0.0772***			
					(0.00123)			
10th grade Math						0.186***		
						(0.00117)		
10th grade Reading							0.0832***	
							(0.00129)	
10th grade Math comparative advantage								0.166***
								(0.00120)
Constant	0.576***	0.489***	0.496***	0.505***	0.475***	0.504***	0.522***	0.468***
	(0.00187)	(0.00578)	(0.00564)	(0.00576)	(0.00572)	(0.00545)	(0.00574)	(0.00551)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	155494	155486	155486	155486	155486	155486	155486	155486
r2	0.028	0.057	0.110	0.070	0.079	0.174	0.081	0.150

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing the STEM track in high school. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 9: Estimation Results for pursuing a STEM Major in University
Conditional on Upper Secondary Enrollment Sample

	(1) STEM in University	(2) STEM in University	(3) STEM in University	(4) STEM in University	(5) STEM in University	(6) STEM in University	(7) STEM in University	(8) STEM in University
Female	-0.161***	-0.156***	-0.138***	-0.160***	-0.143***	-0.125***	-0.170***	-0.114***
	(0.00265)	(0.00266)	(0.00261)	(0.00264)	(0.00269)	(0.00255)	(0.00268)	(0.00263)
6th grade Math			0.0622***					
			(0.00110)					
6th grade Reading				0.0433***				
				(0.00116)				
6th grade Math comparative advantage					0.0325***			
					(0.00111)			
10th grade Math						0.0940***		
						(0.00121)		
10th grade Reading							0.0524***	
							(0.00122)	
10th grade Math comparative advantage								0.0785***
								(0.00119)
Constant	0.237***	0.189***	0.192***	0.202***	0.183***	0.196***	0.209***	0.179***
	(0.00232)	(0.00545)	(0.00536)	(0.00543)	(0.00542)	(0.00528)	(0.00545)	(0.00531)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	75516	75509	75509	75509	75509	75509	75509	75509
r2	0.051	0.086	0.113	0.099	0.093	0.145	0.105	0.127

Note: The sample is restricted to students enrolled in upper secondary education.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: pursuing a standard STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 10: Estimation Results for Pursuing a Broadly-Defined STEM Major in University
Conditional on Upper Secondary Enrollment Sample

	(1) STEM (Broad) in University	(2) STEM (Broad) in University	(3) STEM (Broad) in University	(4) STEM (Broad) in University	(5) STEM (Broad) in University	(6) STEM (Broad) in University	(7) STEM (Broad) in University	(8) STEM (Broad) in University
Female	-0.0903***	-0.0824***	-0.0544***	-0.0878***	-0.0647***	-0.0341***	-0.106***	-0.0203***
	(0.00313)	(0.00312)	(0.00307)	(0.00308)	(0.00317)	(0.00299)	(0.00308)	(0.00312)
6th grade Math			0.0951***					
			(0.00136)					
6th grade Reading				0.0738***				
				(0.00140)				
6th grade Math comparative advantage					0.0433***			
					(0.00137)			
10th grade Math						0.145***		
						(0.00144)		
10th grade Reading							0.0898***	
							(0.00146)	
10th grade Math comparative advantage								0.116***
								(0.00144)
Constant	0.286***	0.204***	0.209***	0.226***	0.196***	0.215***	0.239***	0.189***
	(0.00246)	(0.00638)	(0.00623)	(0.00632)	(0.00634)	(0.00609)	(0.00632)	(0.00618)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	75516	75509	75509	75509	75509	75509	75509	75509
r2	0.011	0.067	0.111	0.093	0.076	0.166	0.106	0.130

Note: The sample is restricted to students enrolled in upper secondary education.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a broadly-defined (including life sciences and medicine) STEM major in university. Controls: Student's birth year, foreign student, status two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 11: Estimation Results for Pursuing a Restrictively-Defined STEM Major in University Conditional on Upper Secondary Enrollment Sample

	(1) STEM (Rest.) in University	(2) STEM (Rest.) in University	(3) STEM (Rest.) in University	(4) STEM (Rest.) in University	(5) STEM (Rest.) in University	(6) STEM (Rest.) in University	(7) STEM (Rest.) in University	(8) STEM (Rest.) in University
Female	-0.159***	-0.155***	-0.139***	-0.158***	-0.144***	-0.127***	-0.168***	-0.117***
	(0.00254)	(0.00254)	(0.00249)	(0.00253)	(0.00256)	(0.00243)	(0.00257)	(0.00249)
6th grade Math			0.0560***					
			(0.00105)					
6th grade Reading				0.0397***				
				(0.00110)				
6th grade Math comparative advantage					0.0286***			
					(0.00106)			
10th grade Math						0.0850***		
						(0.00115)		
10th grade Reading							0.0485***	
							(0.00117)	
10th grade Math comparative advantage								0.0704***
								(0.00113)
Constant	0.219***	0.178***	0.181***	0.189***	0.172***	0.184***	0.197***	0.169***
	(0.00225)	(0.00520)	(0.00513)	(0.00519)	(0.00517)	(0.00507)	(0.00521)	(0.00509)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	75516	75509	75509	75509	75509	75509	75509	75509
r2	0.055	0.086	0.111	0.099	0.093	0.140	0.104	0.124

Note: The sample is restricted to students enrolled in upper secondary education.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a restrictively-defined (math-intensive majors only) STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 12: Estimation Results for pursuing a STEM Track in High School
Conditional on University Enrollment Sample

	(1) STEM in High School	(2) STEM in High School	(3) STEM in High School	(4) STEM in High School	(5) STEM in High School	(6) STEM in High School	(7) STEM in High School	(8) STEM in High School
Female	-0.192***	-0.187***	-0.149***	-0.190***	-0.156***	-0.112***	-0.204***	-0.0897***
	(0.00461)	(0.00470)	(0.00470)	(0.00467)	(0.00479)	(0.00457)	(0.00466)	(0.00474)
6th grade Math			0.111***					
			(0.00227)					
6th grade Reading				0.0539***				
				(0.00233)				
6th grade Math comparative advantage					0.0765***			
					(0.00231)			
10th grade Math						0.182***		
						(0.00214)		
10th grade Reading							0.0699***	
							(0.00234)	
10th grade Math comparative advantage								0.167***
								(0.00221)
Constant	0.647***	0.575***	0.576***	0.590***	0.558***	0.571***	0.599***	0.541***
	(0.00344)	(0.0118)	(0.0115)	(0.0117)	(0.0117)	(0.0111)	(0.0117)	(0.0113)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	45606	45593	45593	45593	45593	45593	45593	45593
r2	0.036	0.077	0.121	0.087	0.098	0.192	0.095	0.172

Note: The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing the STEM track in high school. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 13: Estimation Results for pursuing a STEM Major in University
Conditional on University Enrollment Sample

	(1) STEM in University	(2) STEM in University	(3) STEM in University	(4) STEM in University	(5) STEM in University	(6) STEM in University	(7) STEM in University	(8) STEM in University
Female	-0.292***	-0.292***	-0.270***	-0.293***	-0.271***	-0.249***	-0.299***	-0.233***
	(0.00407)	(0.00416)	(0.00418)	(0.00416)	(0.00423)	(0.00414)	(0.00417)	(0.00424)
6th grade Math			0.0648***					
			(0.00174)					
6th grade Reading				0.0229***				
				(0.00190)				
6th grade Math comparative advantage					0.0519***			
					(0.00182)			
10th grade Math						0.103***		
						(0.00177)		
10th grade Reading							0.0282***	
							(0.00196)	
10th grade Math comparative advantage								0.101***
								(0.00180)
Constant	0.413***	0.375***	0.375***	0.381***	0.363***	0.372***	0.385***	0.354***
	(0.00355)	(0.0101)	(0.00994)	(0.0101)	(0.00999)	(0.00976)	(0.0101)	(0.00978)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	45606	45593	45593	45593	45593	45593	45593	45593
r2	0.113	0.143	0.164	0.146	0.157	0.193	0.147	0.190

Note: The sample is restricted to students enrolled in university.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: pursuing a standard STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 14: Estimation Results for Pursuing a Broadly-Defined STEM Major in University Conditional on University Enrollment Sample

	(1) STEM (Broad) in University	(2) STEM (Broad) in University	(3) STEM (Broad) in University	(4) STEM (Broad) in University	(5) STEM (Broad) in University	(6) STEM (Broad) in University	(7) STEM (Broad) in University	(8) STEM (Broad) in University
Female	-0.188***	-0.185***	-0.152***	-0.187***	-0.158***	-0.119***	-0.201***	-0.101***
	(0.00459)	(0.00470)	(0.00470)	(0.00467)	(0.00478)	(0.00461)	(0.00467)	(0.00477)
6th grade Math			0.0967***					
			(0.00215)					
6th grade Reading				0.0497***				
				(0.00225)				
6th grade Math comparative advantage					0.0648***			
					(0.00219)			
10th grade Math						0.159***		
						(0.00211)		
10th grade Reading							0.0635***	
							(0.00228)	
10th grade Math comparative advantage								0.145***
								(0.00214)
Constant	0.499***	0.427***	0.428***	0.441***	0.413***	0.424***	0.449***	0.398***
	(0.00360)	(0.0115)	(0.0113)	(0.0115)	(0.0114)	(0.0110)	(0.0115)	(0.0111)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	45606	45593	45593	45593	45593	45593	45593	45593
r2	0.036	0.073	0.109	0.083	0.089	0.165	0.088	0.147

Note: The sample is restricted to students enrolled in university.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a broadly-defined (including life sciences and medicine) STEM major in university. Controls: Student's birth year, foreign student, status two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 15: Estimation Results for Pursuing a Restrictively-Defined STEM Major in University Conditional on University Enrollment Sample

	(1) STEM (Rest.) in University	(2) STEM (Rest.) in University	(3) STEM (Rest.) in University	(4) STEM (Rest.) in University	(5) STEM (Rest.) in University	(6) STEM (Rest.) in University	(7) STEM (Rest.) in University	(8) STEM (Rest.) in University
Female	-0.287***	-0.287***	-0.267***	-0.288***	-0.268***	-0.248***	-0.294***	-0.234***
	(0.00394)	(0.00403)	(0.00403)	(0.00402)	(0.00408)	(0.00399)	(0.00403)	(0.00407)
6th grade Math			0.0590***					
			(0.00166)					
6th grade Reading				0.0225***				
				(0.00181)				
6th grade Math comparative advantage					0.0459***			
					(0.00174)			
10th grade Math						0.0940***		
						(0.00169)		
10th grade Reading							0.0283***	
							(0.00188)	
10th grade Math comparative advantage								0.0910***
								(0.00172)
Constant	0.382***	0.350***	0.351***	0.357***	0.340***	0.348***	0.360***	0.332***
	(0.00350)	(0.00972)	(0.00960)	(0.00972)	(0.00964)	(0.00946)	(0.00973)	(0.00948)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	45606	45593	45593	45593	45593	45593	45593	45593
r2	0.118	0.147	0.166	0.150	0.159	0.193	0.152	0.189

Note: The sample is restricted to students enrolled in university.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing a restrictively-defined (math-intensive majors only) STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 16: Estimation Results for Pursuing a STEM Track in High School by Math Performance Quartile
Conditional on University Enrollment Sample

	1st Quartile (Lowest)		**2nd Quartile**		**3rd Quartile**		**4th Quartile (Highest)**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	STEM in High School	STEM in High School	STEM in High School	STEM in High School	STEM in High School	STEM in High School	STEM in High School	STEM in High School
Female	-0.0839***	-0.0688***	-0.117***	-0.104***	-0.121***	-0.112***	-0.106***	-0.0984***
	(0.0102)	(0.0104)	(0.0101)	(0.0104)	(0.00948)	(0.00987)	(0.00824)	(0.00847)
10th grade Math	0.112***		0.223***		0.238***		0.195***	
	(0.00651)		(0.0195)		(0.0261)		(0.0241)	
10th grade Math comparative advantage		0.0811***		0.0721***		0.0542***		0.0540***
		(0.00560)		(0.00961)		(0.0103)		(0.00987)
Constant	0.482***	0.409***	0.568***	0.520***	0.512***	0.599***	0.583***	0.735***
	(0.0238)	(0.0228)	(0.0251)	(0.0249)	(0.0281)	(0.0257)	(0.0334)	(0.0237)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	11360	11360	11401	11401	11301	11301	11329	11329
r2	0.142	0.137	0.129	0.123	0.137	0.133	0.125	0.122

Note: The sample is restricted to students enrolled in university.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: Pursuing the STEM track in high school. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

Table 17: Estimation Results for Pursuing a STEM Major in University by Math Performance Quartile
Conditional on University Enrollment Sample

	1st Quartile (Lowest)		**2nd Quartile**		**3rd Quartile**		**4th Quartile (Highest)**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	STEM in University	STEM in University	STEM in University	STEM in University	STEM in University	STEM in University	STEM in University	STEM in University
Female	-0.130***	-0.120***	-0.201***	-0.190***	-0.255***	-0.245***	-0.279***	-0.269***
	(0.00765)	(0.00769)	(0.00847)	(0.00872)	(0.00908)	(0.00944)	(0.00950)	(0.00980)
10th grade Math	0.0468***		0.154***		0.191***		0.234***	
	(0.00407)		(0.0154)		(0.0247)		(0.0285)	
10th grade Math comparative advantage		0.0414***		0.0582***		0.0522***		0.0713***
		(0.00364)		(0.00772)		(0.00987)		(0.0116)
Constant	0.235***	0.212***	0.348***	0.315***	0.331***	0.395***	0.350***	0.525***
	(0.0166)	(0.0158)	(0.0207)	(0.0205)	(0.0265)	(0.0240)	(0.0395)	(0.0280)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	11360	11360	11401	11401	11301	11301	11329	11329
r2	0.154	0.155	0.168	0.165	0.181	0.179	0.180	0.177

Note: The sample is restricted to students enrolled in university.

Estimates include school and cohort fixed effects. Standard errors are clustered at the school level. Outcome: pursuing a standard STEM major in university. Controls: Student's birth year, foreign student status, two-member household, same-sex household, household with both tutors holding a college degree, household with only one tutor holding a college degree, household with only one tutor born abroad, and household with both tutors born abroad.

All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one.

Stars indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses.

Source: Own elaboration based on IDESCAT data.

References

Bachman, J. G., O'Malley, P. M., Freedman-Doan, P., Trzesniewski, K. H., and Donnellan, M. B. (2011). Adolescent self-esteem: Differences by race/ethnicity, gender, and age. Self and Identity, 10:445–473.

Bordalo, P., Coffman, K. B., Gennaioli, N., and Shleifer, A. (2016). Beliefs about gender. Technical report.

Bottazzi, L. and Lusardi, A. (2020). Nber working paper series stereotypes in financial literacy: Evidence from pisa stereotypes in financial literacy: Evidence from pisa. Technical report.

Breda, T., Jouini, E., and Napp, C. (2023). Gender differences in the intention to study math increase with math performance. Nature Communications, 14.

Breda, T. and Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. Proceedings of the National Academy of Sciences of the United States of America, 116:15435–15440.

Bursztyn, L. and Yang, D. Y. (2022). Misperceptions about others.

Fry, B. R., Kennedy, B., and Funk, C. (2021). Stem jobs see uneven progress in increasing gender, racial and ethnic diversity higher education pipeline suggests long path ahead for increasing diversity, especially in fields like computing and engineering for media or other inquiries. Technical report.

Goulas, S., Griselda, S., and Megalokonomou, R. (2020). Comparative advantage and gender gap in stem. Technical report.

Guadalupe, L. D. C. M. (2018). More women in tech? evidence from a field experiment addressing social identity. Technical report.

Kahn, S. and Ginther, D. (2017). Women and stem. Technical report.

Stoet, G. and Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of pisa data. PLoS ONE, 8.

Wang, M. T., Eccles, J. S., and Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. Psychological Science, 24:770–775.

World Bank (2025). Educational attainment, at least bachelor's or equivalent, population 25+, total (%) (cumulative). `https://data.worldbank.org/indicator/SE.TER.CUAT.BA.ZS`. Accessed: 2025-05-07.

Xie, F., Xin, Z., Chen, X., and Zhang, L. (2019). Gender difference of chinese high school students' math anxiety: The effects of self-esteem, test anxiety and general anxiety. Sex Roles, 81:235–244.

Appendix - Figures

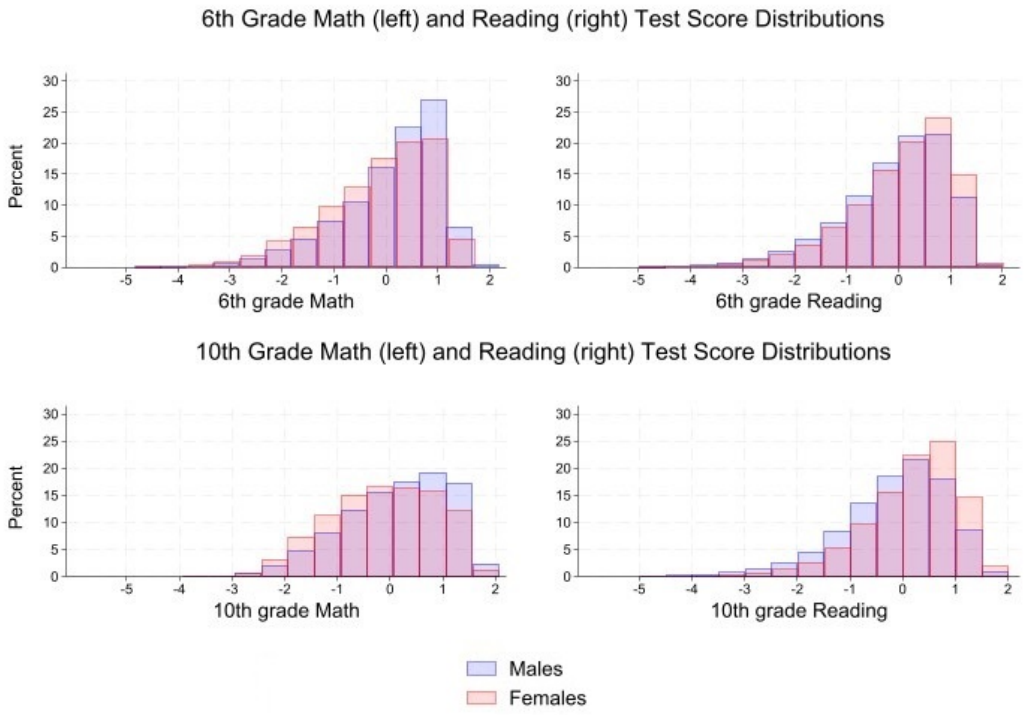


Figure A.1: **Gender Gap in Academic Performance: Histogram Distributions of 6^{th} and 10^{th} Grade Test Scores.** These graphs display the distribution of standardized test scores (mean 0, standard deviation 1) in math (left column) and reading (right column) for male (blue) and female (red) students. The top row corresponds to 6^{th} grade, and the bottom row to 10^{th} grade. The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment. Source: Own elaboration based on IDESCAT data.

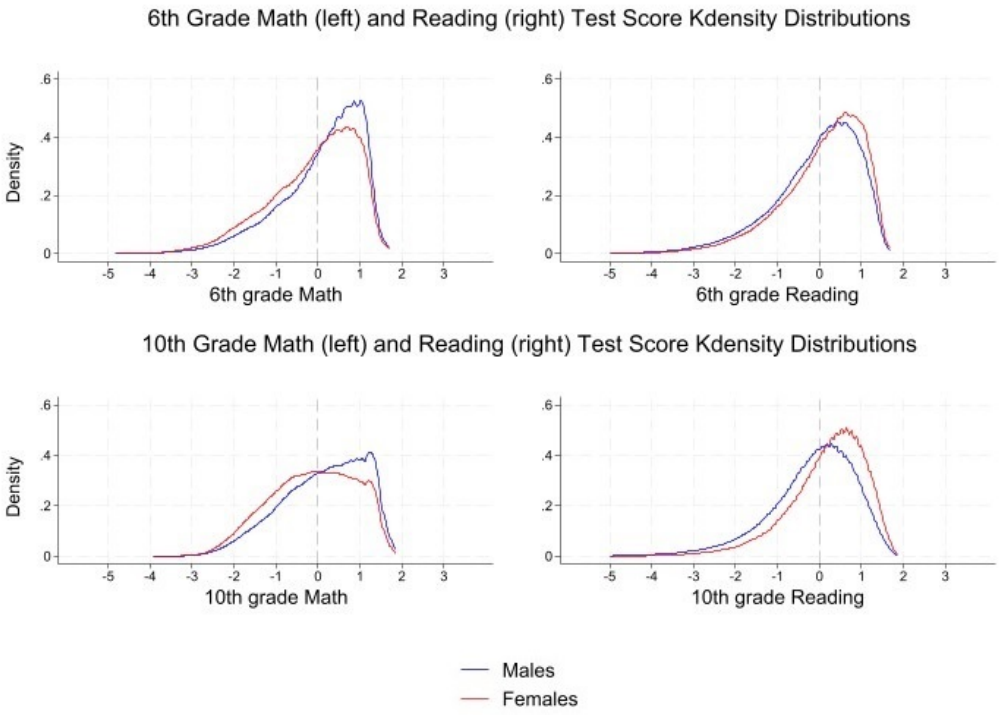


Figure A.2: **Gender Gap in Academic Performance: Kernel Density Estimates of 6^{th} and 10^{th} Grade Test Scores.** These graphs display kernel density estimates of standardized test scores (mean 0, standard deviation 1) in math (left column) and reading (right column) for male (blue) and female (red) students. The top row corresponds to 6^{th} grade grade, and the bottom row to 10^{th} grade grade. The sample is representative of 10^{th} grade students in Catalonia and is unconditional on post-compulsory education enrollment. Source: Own elaboration based on IDESCAT data.

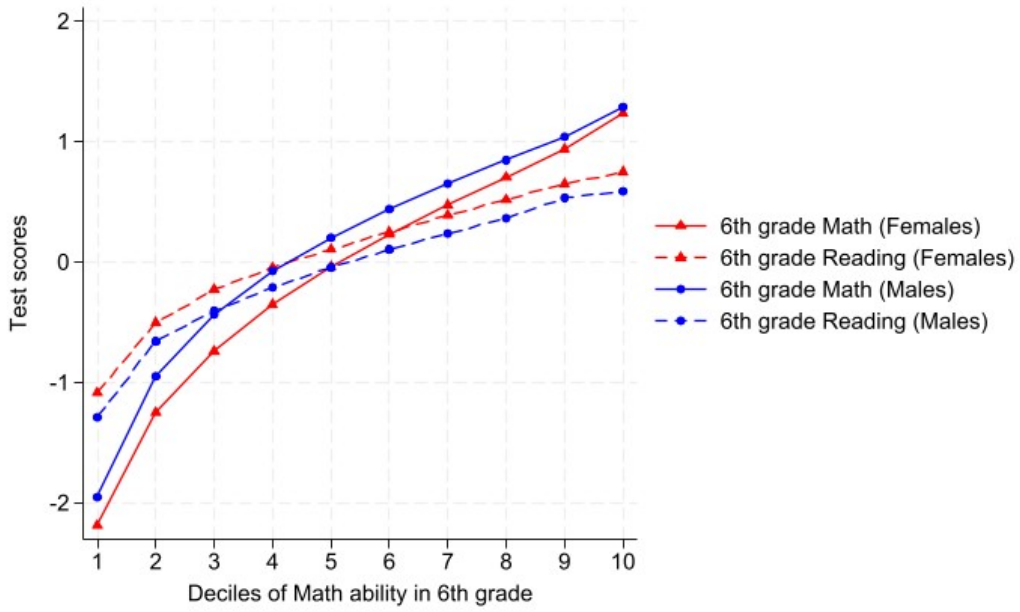


Figure A.3: **Comparative Patterns of Math and Reading Performance in 6th Grade, by Gender.** The figure displays average standardized scores (mean 0, standard deviation 1) in math and reading across deciles of 6th-grade math ability, separately for male and female students. For both genders, higher math performance is systematically associated with higher reading performance. The sample is representative of 10th grade students in Catalonia and is unconditional on post-compulsory education enrollment. Source: Own elaboration based on IDESCAT data.

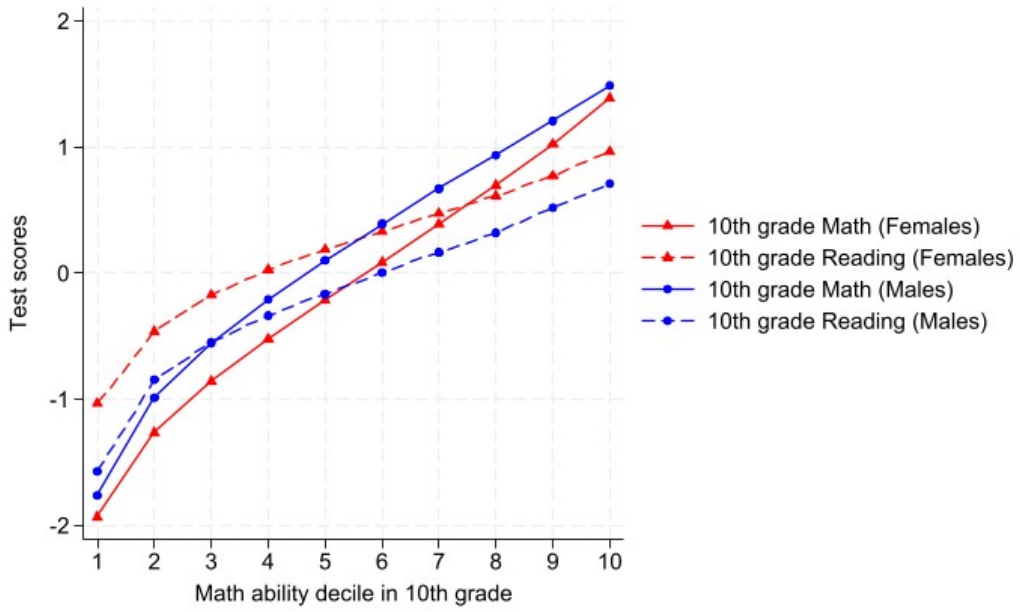


Figure A.4: **Comparative Patterns of Math and Reading Performance in 10th Grade, by Gender.** The figure displays average standardized test scores (mean 0, standard deviation 1) in math and reading across deciles of 10th-grade math ability, separately by gender. While math and reading performance remain positively associated, gender differences in achievement are more pronounced than in 6th grade. The sample is representative of 10th grade students in Catalonia and is unconditional on post-compulsory education enrollment. Source: Own elaboration based on IDESCAT data.

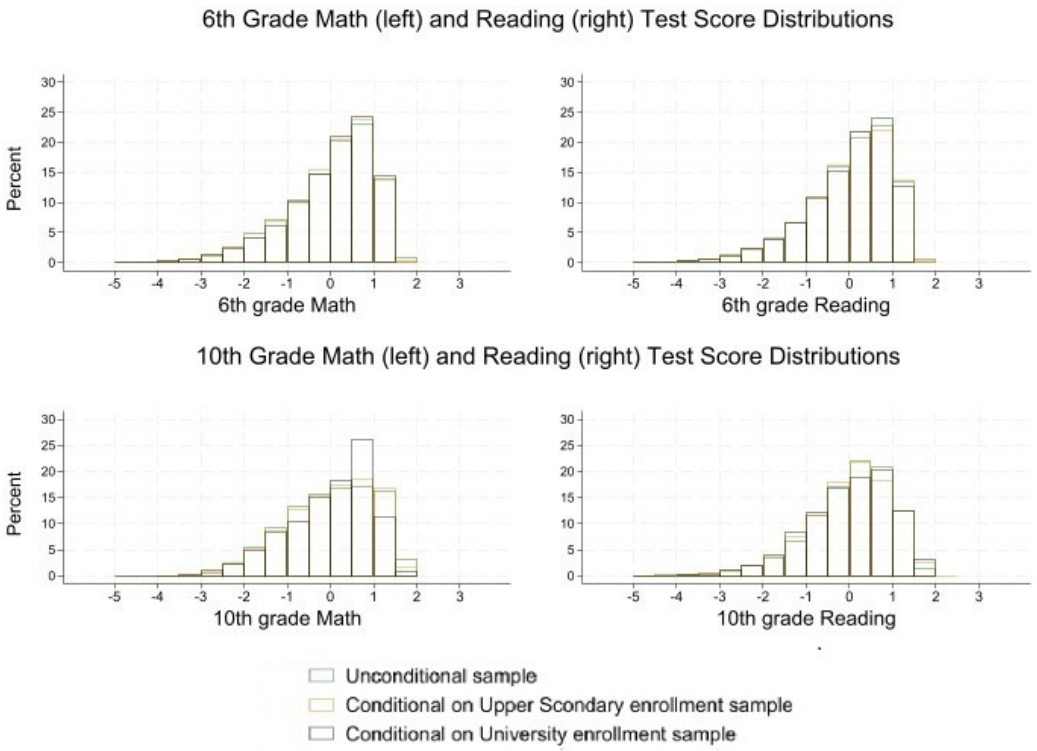


Figure A.5: **Changes in Sample Composition: Histogram Distributions of 6^{th} and 10^{th} Grade Test Scores.** These graphs display the distribution of standardized test scores (mean 0, standard deviation 1) in math (left column) and reading (right column) for three distinct samples: the unconditional sample (green), the sample conditional on high school enrollment (orange), and the sample conditional on university enrollment (black). The figure highlights how the distribution of academic performance shifts across samples that represent progressively more selective educational pathways. The top row corresponds to 6^{th} grade, and the bottom row to 10^{th} grade. Source: Own elaboration based on IDESCAT data.

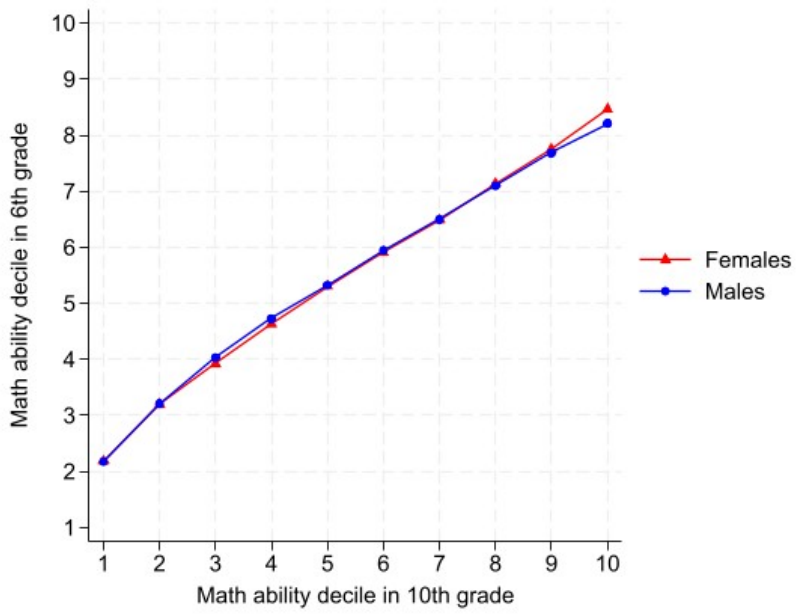


Figure A.6: **Progression of Math Ability from 6th to 10th Grade, by Gender.** The figure plots average 10th-grade math ability deciles against 6th-grade math ability deciles, separately for female and male students. The relationship is nearly linear, indicating strong persistence of relative performance in math over time. Differences between boys and girls are minimal across the distribution, suggesting that math trajectories remain broadly parallel by gender. Source: Own elaboration based on IDESCAT data.

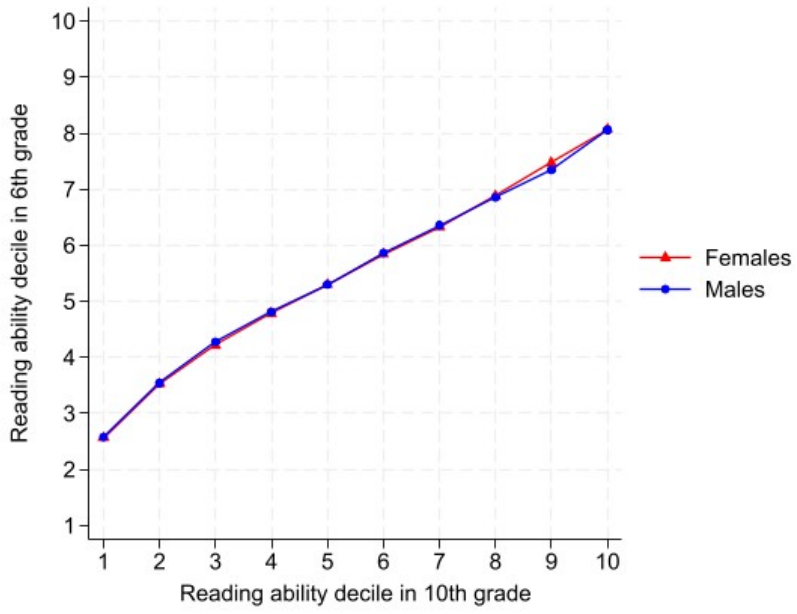


Figure A.7: **Progression of Reading Ability from 10th to 6th Grade, by Gender.** The figure plots average 6th-grade reading ability deciles against 10th-grade reading ability deciles, separately for female and male students. The relationship is close to linear, reflecting strong persistence in relative reading performance over time. Gender differences are negligible across the distribution, suggesting parallel trajectories for boys and girls in reading achievement. Source: Own elaboration based on IDESCAT data.

Appendix - Tables

Table A.1: University-Level STEM Classification Based on Field-of-Study Codes

Field of Study	Restrictive STEM Definition	Standard STEM Definition	Broad STEM Definition
Physics	✓	✓	✓
Mathematics	✓	✓	✓
Statistics	✓	✓	✓
Chemistry	✓	✓	✓
ICTs	✓	✓	✓
Software development and analysis	✓	✓	✓
Electronics and automation	✓	✓	✓
Mechanics and metal trades	✓	✓	✓
Motor vehicles, ships and aircraft	✓	✓	✓
Engineering	✓	✓	✓
Construction and civil engineering	✓	✓	✓
Architecture and urban planning	✗	✓	✓
Environmental sciences	✗	✓	✓
Earth sciences	✗	✓	✓
Computer use	✗	✓	✓
Electricity and energy	✗	✓	✓
Biology	✗	✗	✓
Biochemistry	✗	✗	✓
Agriculture	✗	✗	✓
Veterinary	✗	✗	✓
Medicine	✗	✗	✓
Dentistry	✗	✗	✓
Nursing and midwifery	✗	✗	✓
Therapy and rehabilitation	✗	✗	✓
Pharmacy	✗	✗	✓
Health, not elsewhere classified	✗	✗	✓
Forestry	✗	✗	✓

Note: A checkmark indicates that the corresponding field of study is included in the specified STEM classification. Classifications are based on university field-of-study codes from IDESCAT. The restrictive STEM definition comprises only math-intensive fields and core physical sciences, excluding life sciences and health-related disciplines. The standard STEM definition additionally includes environmental sciences, earth sciences, and computer use, while the broad STEM definition further incorporates life sciences, health-related fields, and agriculture. Source: Own elaboration based on IDESCAT data.

Table A.2: Descriptive Statistics
Conditional on Upper Secondary Enrollment Sample

Variable	Mean	SD	Min	Max	Obs.
Female	0.55	0.50	0.00	1.00	155494
6th grade Math	0.00	1.00	-5.65	1.61	155494
6th grade Reading	0.00	1.00	-7.82	1.65	155494
6th grade Math comparative advantage	0.00	1.00	-7.08	5.57	155494
10th grade Math	0.00	1.00	-4.76	1.76	155494
10th grade Reading	0.00	1.00	-11.05	2.01	155494
10th grade Math comparative advantage	0.00	1.00	-5.84	6.50	155494
Birth year	2003.53	2.27	1998.00	2007.00	155494
Foreign student	0.06	0.24	0.00	1.00	155494
Two-member household	0.91	0.28	0.00	1.00	155494
Same-sex household	0.01	0.11	0.00	1.00	155494
HH with both tutors college-educated	0.24	0.43	0.00	1.00	155494
HH with only one tutor college-educated	0.25	0.43	0.00	1.00	155494
HH with only one tutor foreign-born	0.12	0.33	0.00	1.00	155494
HH with both tutors foreign-born	0.10	0.30	0.00	1.00	155494
STEM in high school	0.48	0.50	0.00	1.00	155494
STEM in university	0.15	0.35	0.00	1.00	75516
STEM (w. life sciences and medicine) in university	0.24	0.42	0.00	1.00	75516
STEM (math-intensive majors only) in university	0.13	0.34	0.00	1.00	75516

Note: The sample is conditional on students enrolling in upper secondary education. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are only observed for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table A.3: Descriptive Statistics
Conditional on University Enrollment Sample

Variable	Mean	SD	Min	Max	Obs.
Female	0.58	0.49	0.00	1.00	45606
6th grade Math	0.00	1.00	-6.50	1.34	45606
6th grade Reading	0.00	1.00	-9.57	1.47	45606
6th grade Math comparative advantage	0.00	1.00	-7.87	6.15	45606
10th grade Math	0.00	1.00	-5.58	1.59	45606
10th grade Reading	0.00	1.00	-11.33	1.95	45606
10th grade Math comparative advantage	0.00	1.00	-4.63	6.23	45606
Birth year	2001.45	1.12	2000.00	2004.00	45606
Foreign student	0.05	0.22	0.00	1.00	45606
Two-member household	0.91	0.29	0.00	1.00	45606
Same-sex household	0.01	0.10	0.00	1.00	45606
HH with both tutors college-educated	0.29	0.45	0.00	1.00	45606
HH with only one tutor college-educated	0.26	0.44	0.00	1.00	45606
HH with only one tutor foreign-born	0.13	0.33	0.00	1.00	45606
HH with both tutors foreign-born	0.06	0.23	0.00	1.00	45606
STEM in high school	0.54	0.50	0.00	1.00	45606
STEM in university	0.24	0.43	0.00	1.00	45606
STEM (w. life sciences and medicine) in university	0.39	0.49	0.00	1.00	45606
STEM (math-intensive majors only) in university	0.22	0.41	0.00	1.00	45606

Note: The sample is conditional on students enrolling in university. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. Source: Own elaboration based on IDESCAT data.

Table A.4: Statistical Differences
Unconditional Sample vs. Sample Conditional on Upper Secondary Enrollment

Variable	Diff./SE	Unconditional sample	Cond. on Upper Secondary sample	p-value
Female	0.05***	0.51	0.55	0.00
	(0.00)			
6th grade Math	0.00	0.00	0.00	1.00
	(0.00)			
6th grade Reading	0.00	0.00	0.00	1.00
	(0.00)			
6th grade math comparative advantage	0.00	0.00	0.00	1.00
	(0.00)			
10th grade Math	0.00	0.00	0.00	1.00
	(0.00)			
10th grade Reading	0.00	0.00	0.00	1.00
	(0.00)			
10th grade math comparative advantage	0.00	0.00	0.00	1.00
	(0.00)			
Birth year	-0.05***	2003.58	2003.53	0.00
	(0.01)			
Foreign student	-0.01***	0.07	0.06	0.00
	(0.00)			
Two-member household	0.01***	0.90	0.91	0.00
	(0.00)			
Same-sex household	0.00	0.01	0.01	0.42
	(0.00)			
HH with both tutors college-educated	0.05***	0.19	0.24	0.00
	(0.00)			
HH with only one tutor college-educated	0.03***	0.22	0.25	0.00
	(0.00)			
HH with only one tutor foreign-born	-0.01***	0.13	0.12	0.00
	(0.00)			
HH with both tutors foreign-born	-0.01***	0.11	0.10	0.00
	(0.00)			
STEM in high school	0.14***	0.34	0.48	0.00
	(0.00)			
Observations	375131			
STEM in university	0.04***	0.11	0.15	0.00
	(0.00)			
STEM (w. life sciences and medicine) in university	0.06***	0.17	0.24	0.00
	(0.00)			
STEM (math-intensive majors only) in university	0.03***	0.10	0.13	0.00
	(0.00)			
Observations	179467			

Note: Column "Diff./SE" reports the difference in means between the unconditional sample and the sample conditional on upper secondary enrollment, with standard errors in parentheses. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are available exclusively for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table A.5: Statistical Differences
Conditional on Upper Secondary Enrollment vs. Conditional on University Enrollment

Variable	Diff./SE	Cond. on Upper Secondary sample	Cond. on University sample	p-value
Female	0.03***	0.55	0.58	0.00
	(0.00)			
6th grade Math	0.00	0.00	0.00	1.00
	(0.01)			
6th grade Reading	0.00	0.00	0.00	1.00
	(0.01)			
6th grade Math comparative advantage	0.00	0.00	0.00	1.00
	(0.01)			
10th grade Math	0.00	0.00	0.00	1.00
	(0.01)			
10th grade Reading	0.00	0.00	0.00	1.00
	(0.01)			
10th grade Math comparative advantage	0.00	0.00	0.00	1.00
	(0.01)			
Birth year	-2.08***	2003.53	2001.45	0.00
	(0.01)			
Foreign student	-0.01***	0.06	0.05	0.00
	(0.00)			
Two-member household	0.00***	0.91	0.91	0.00
	(0.00)			
Same-sex household	0.00***	0.01	0.01	0.00
	(0.00)			
HH with both tutors college-educated	0.05***	0.24	0.29	0.00
	(0.00)			
HH with only one tutor college-educated	0.01***	0.25	0.26	0.00
	(0.00)			
HH with only one tutor foreign-born	0.00	0.12	0.13	0.19
	(0.00)			
HH with both tutors foreign-born	-0.05***	0.10	0.06	0.00
	(0.00)			
STEM in high school	0.05***	0.48	0.54	0.00
	(0.00)			
Observations	201100			
STEM in university	0.10***	0.15	0.24	0.00
	(0.00)			
STEM (w. life sciences and medicine) in university	0.15***	0.24	0.39	0.00
	(0.00)			
STEM (math-intensive majors only) in university	0.09***	0.13	0.22	0.00
	(0.00)			
Observations	121122			

Note: Column "Diff./SE" reports the difference in means between the sample conditional on high school enrollment and the sample conditional on university enrollment, with standard errors in parentheses. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are available exclusively for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table A.6: Descriptive Statistics by Gender
Conditional on Upper Secondary Enrollment Sample

Variable	Diff./SE	Females	Males	p-value
6th grade Math	0.31***	-0.14	0.17	0.00
	(0.01)			
6th grade Reading	-0.07***	0.03	-0.04	0.00
	(0.01)			
6th grade Math comparative advantage	0.41***	-0.19	0.23	0.00
	(0.00)			
10th grade Math	0.40***	-0.18	0.22	0.00
	(0.00)			
10th grade Reading	-0.27***	0.12	-0.15	0.00
	(0.01)			
10th grade Math comparative advantage	0.60***	-0.27	0.33	0.00
	(0.00)			
Birth year	0.03**	2003.52	2003.55	0.02
	(0.01)			
Foreign student	-0.01***	0.07	0.05	0.00
	(0.00)			
Two-member household	0.01***	0.91	0.92	0.00
	(0.00)			
Same-sex household	0.00	0.01	0.01	0.20
	(0.00)			
HH with both tutors college-educated	0.04***	0.22	0.26	0.00
	(0.00)			
HH with only one tutor college-educated	0.02***	0.24	0.26	0.00
	(0.00)			
HH with only one tutor foreign-born	-0.01***	0.13	0.12	0.00
	(0.00)			
HH with both tutors foreign-born	-0.01***	0.11	0.09	0.00
	(0.00)			
STEM in high school	0.17***	0.41	0.58	0.00
	(0.00)			
Observations	155494			
STEM in university	0.16***	0.08	0.24	0.00
	(0.00)			
STEM (w. life sciences and medicine) in university	0.09***	0.20	0.29	0.00
	(0.00)			
STEM (math-intensive majors only) in university	0.16***	0.06	0.22	0.00
	(0.00)			
Observations	75516			

Note: The sample is conditional on students enrolling in high school. Column "Diff./SE" reports the difference in means between females and males, with standard errors in parentheses. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. University outcomes are available exclusively for cohorts born between 2000 and 2003. Source: Own elaboration based on IDESCAT data.

Table A.7: Descriptive Statistics by Gender
Conditional on University Enrollment Sample

Variable	Diff./SE	Females	Males	p-value
6th grade Math	0.36***	-0.15	0.21	0.00
	(0.01)			
6th grade Reading	-0.03***	0.01	-0.01	0.01
	(0.01)			
6th grade Math comparative advantage	0.42***	-0.18	0.24	0.00
	(0.01)			
10th grade Math	0.43***	-0.18	0.25	0.00
	(0.01)			
10th grade Reading	-0.23***	0.10	-0.13	0.00
	(0.01)			
10th grade Math comparative advantage	0.59***	-0.25	0.34	0.00
	(0.01)			
Birth year	0.02*	2001.44	2001.46	0.09
	(0.01)			
Foreign student	-0.02***	0.06	0.04	0.00
	(0.00)			
Two-member household	0.01***	0.91	0.92	0.00
	(0.00)			
Same-sex household	0.00	0.01	0.01	0.70
	(0.00)			
HH with both tutors college-educated	0.06***	0.26	0.33	0.00
	(0.00)			
HH with only one tutor college-educated	0.01**	0.26	0.27	0.01
	(0.00)			
HH with only one tutor foreign-born	-0.01**	0.13	0.12	0.04
	(0.00)			
HH with both tutors foreign-born	-0.01***	0.06	0.05	0.00
	(0.00)			
STEM in high school	0.19***	0.46	0.65	0.00
	(0.00)			
STEM in university	0.29***	0.12	0.41	0.00
	(0.00)			
STEM (w. life sciences and medicine) in university	0.19***	0.31	0.50	0.00
	(0.00)			
STEM (math-intensive majors only) in university	0.29***	0.10	0.38	0.00
	(0.00)			
Observations	45606			

Note: The sample is conditional on students enrolling in university. Column "Diff./SE" reports the difference in means between females and males, with standard errors in parentheses. Stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All scores (math, reading, and math comparative advantage) are standardized by cohort and subject to have a mean of zero and a standard deviation of one. Source: Own elaboration based on IDESCAT data.