Available online at www.sciencedirect.com

ScienceDirect

Journal of Policy Modeling xxx (xxxx) xxx–xxx

www.elsevier.com/locate/jpm

Decoding girls' STEM high school choices: Ability, confidence, stereotypes

Dominique Cappelletti [a,b], M. Vittoria Levati [a,*], Matteo Ploner [c]

[a] Department of Economics, University of Verona, Via Cantarane 24, Verona 37129, Italy
[b] FBK-IRVAPP, Via S. Croce 77, Trento 38122, Italy
[c] Department of Economics & Management, University of Trento, Via Inama 5, Trento 38122, Italy

Received 12 December 2025; Received in revised form 19 March 2026; Accepted 25 March 2026
Available online xxxx

Abstract

The underrepresentation of females in STEM fields hinders productivity and perpetuates labor market inequalities. In countries where children are tracked into educational trajectories from high school (as in Italy, 8th grade), it is crucial to understand what drives gendered pathways before educational segregation starts. Collecting experimental and survey data from Italian 8th graders, we find that perceived advantageous comparisons with peers in math ability and counter-stereotypical beliefs increase the likelihood that girls enroll in a math-intensive track during high school. Policy initiatives improving girls' expectations about their relative math performance may thus encourage female students to pursue STEM tracks.

JEL Classification: C93; J16; J24; I24

Keywords: High school enrollment; Math ability; Gender stereotypes; Beliefs

* Corresponding author.
E-mail address: vittoria.levati@univr.it (M. Vittoria Levati).

https://doi.org/10.1016/j.jpolmod.2026.107046
0161-8938/

Please cite this article as: D. Cappelletti, M. Vittoria Levati and M. Ploner, Decoding girls' STEM high school choices: Ability, confidence, stereotypes, Journal of Policy Modeling, https://doi.org/10.1016/j.jpolmod.2026.107046

1. Introduction

Skills in science, technology, engineering, and mathematics (STEM) are increasingly important in today's innovation-driven global economy. Employers seek workers with STEM skills and knowledge to foster technological advancements and enhance productivity in all sectors. Despite the promising job prospects, the growing demand for STEM skills is not matched by a corresponding increase in the number of youngsters pursuing studies in these fields. The problem is particularly pronounced among female students, who remain significantly underrepresented in math- and science-intensive educational tracks.[1]

The marked gender differences in educational choices have profound economic and social consequences. Reducing the gender gap in STEM education could, for instance, increase the labor supply in high value-added sectors, potentially rising the European Union's per capita GDP by up to 0.9 % within a decade and by up to 3 % over three decades (EIGE, 2017). Moreover, gendered field-of-study choices contribute significantly to labor market inequality as STEM graduates earn, on average, more than humanities majors (Kirkeboen et al., 2016) and several studies provide evidence that a substantial part of the gender wage gap is attributable to gender differences in college and university majors (e.g., Blau and Kahn, 2017; Francesconi and Parey, 2018; Card and Payne, 2021). It is therefore not surprising that the underrepresentation of women in STEM education has become a major concern for policy-makers and researchers, spurring efforts to identify the factors that discourage female participation in math- and science-intensive studies. While progress has been made, there remains a need for research that links experimentally measured individual characteristics of male and female students to their educational choices.

In this paper, we make a step in this direction by conducting a lab-in-the-field experiment with 534 Italian 8th graders. In the experiment, we assess a number of gendered math-related factors and complement the experimental data with an in-depth survey asking the students about the high (or upper secondary) school track they enrolled in, among other pieces of information.

Italy's educational system, characterized by early tracking, provides a relevant context for our study. During grade 8 (the final year of middle school), typically at age 14, Italian students formally choose the type of high school track they want to enroll in, where tracks vary in the levels of math and science intensity. The high school track choice is hard to reverse and greatly affects tertiary education decisions and, in turn, labor market opportunities. Data from the Italian Ministry of Education show that most of the students attending math- and science-intensive high school curricula enroll in STEM universities, and Granato (2023) estimates that about half of the gender gap in STEM graduation in Italy is attributable to gendered high school choices. Thus, in educational systems characterized by early tracking (like the Italian one), understanding the factors that contribute to steer girls away from math- and science-intensive tracks is crucial for designing effective interventions.

We focus on three potential such factors: gender differences in math ability, gender differences in beliefs about math ability, and individual math-gender stereotypes. Prior research shows that girls underperform boys in math standardized tests at all school grades (Fryer and Levitt, 2010; OECD, 2015; Contini et al., 2017), particularly at the upper and lower tails of the math performance distribution (Pope and Sydnor, 2010; Matteucci and Mignani, 2021). However, the gender gap in math test scores may exaggerate the actual differences in math

[1] See Eurostat Tertiary Education Statistics at link.

skills since economic experiments indicate that males and females respond differently to competitive test-taking environments (Niederle and Vesterlund, 2010; Carpenter et al., 2018), with females, in particular, tending to avoid competing against males (Buser et al., 2023; Paryavi et al., 2025). Moreover, several meta-analyses report no or minimal gender differences in math ability among primary, middle, and high school students (Hyde et al., 2008; Else-Quest et al., 2010; Lindberg et al., 2010; Stoet and Geary, 2018).

Self-perception of – or confidence in – math ability is another factor that may explain gendered educational choices. Murphy and Weinhardt (2020) provide evidence that students who rank higher in a given subject during primary school tend to exhibit greater confidence in that subject later on, and this increased confidence is, in turn, associated with a higher likelihood of selecting math-related courses at the end of secondary education. In addition, a number of studies reveal that girls are less confident in STEM-related subjects than boys (Dahlbom et al., 2011; Shi, 2018; Exley and Kessler, 2022).

A third factor that has received increasing attention is math-gender stereotypes or biases, i.e., the preconception that females have less math ability than males. Traditional gender stereotypes, held by females and/or embedded in the family and school environment, can impair girls' math performance (Flore and Wicherts, 2015), negatively affect their math confidence (Muzzatti and Agnoli, 2007; Passolunghi et al., 2014), and shape their interest in math related subjects (Plante et al., 2019). Recent economic research has been devoted to the study of teachers' gender stereotypes and their short- and long-run consequences on performance and educational choice (Breda and Ly, 2015; Breda and Hillion, 2016; Alan et al., 2018; Lavy and Sand, 2018; Carlana, 2019). There are also economic lab experiments that show how gender stereotypes influence beliefs and decisions in male-typed domains (Reuben et al., 2014; Bordalo et al., 2019). However, existing research has not yet linked experimentally measured individual math-gender stereotypes to high school track choices – a gap this paper seeks to address.

In our lab-in-the-field experiment, 8th graders from five public middle schools in Trento (northeastern Italy) perform an incentivized math task and we use performance in this task as a proxy for their math ability. Based on this performance, we elicit incentivized beliefs about own absolute math ability and own relative-to-the-median math ability, which are used to construct measures of confidence. Performance in the math task serves also as the basis for our incentivized measure of individual gender stereotypes, which is collected through a belief elicitation task requiring participants to guess the average math performance of a male and female student. To gain a comprehensive picture of the stereotypes at play, we also administer more traditional measures of stereotypes, namely the gender-science Implicit Association Test (IAT) and explicit questions asking for consciously-held beliefs about gender differences in math.[2] These experimentally derived variables are supplemented with survey responses from students. The survey gathers data on their high school track choices, their teachers' non-binding recommendations, their perceived value of math, and the interest in math conveyed by the school.

We document a significant gender gap in reported high school enrollment. Notably, highly math-talented boys are more likely than equally talented girls to report enrollment in a math-intensive high school track. We do not identify a consistent and systematic math-gender

[2] Economists have recently studied gender stereotypes using either implicit (Reuben et al., 2014; Carlana, 2019) or explicit (Alan et al., 2018) measures, and there are a few studies in psychology that investigate math-gender stereotypes at both the implicit and explicit levels in the same children sample (Steffens et al., 2010; Galdi et al., 2014; Passolunghi et al., 2014).

stereotype in the population under study. In fact, all participants, regardless of gender, believe that females on average outperform males in the math task. As to the factors related to the decision to enroll in a math-intensive high school, we observe that they differ by gender. Among boys, enrollment in a math-intensive track is positively associated with beliefs about own absolute math ability as well as with implicit and explicit math-gender stereotypes. Among girls, by contrast, enrollment is more strongly associated with perceived comparisons with peers and with gender counter-stereotypical beliefs about math.

Our paper adds to the economics literature in two main ways. First, this is one of the few papers examining female enrollment in upper secondary education. Most previous studies have focused on enrollment and persistence in STEM tertiary education, while gender segregation in educational pathways may emerge earlier. In many European and Asian countries, students are tracked into math-intensive versus non math-intensive routes in high school, when they are 15 years old or earlier. Yet, apart from a few studies, such as Buser et al. (2014) for the Netherlands, Buser et al. (2017) for Switzerland, Rapoport and Thibout (2018) for France, and Mouganie and Wang (2020) for China, little is known about what drives girls' high school track enrollment.

Second, we contribute to the body of research that employs an experimental approach to understand gender differences in educational choices, in particular the one that links individual characteristics measured in the laboratory with choices outside the laboratory. Most prominently, Buser et al. (2014) and Buser et al. (2017) measure students' competitiveness in an experiment and show that it significantly predicts the selection of math-intensive high school courses and that the gender difference in competitiveness partly explains the gender difference in course selection. Reuben et al. (2017) find that experimentally measured competitiveness and overconfidence are positively related to expected future earnings, which are in turn related to college major choices.

The remainder of the article is organized as follows. Section 2 provides details of the institutional context and the alternative educational options Italian students have at the end of middle school. Section 3 describes the procedure of data collection and explains the experimental and survey variables. Section 4 contains the results. Section 5 offers concluding remarks.

2. Institutional setting

In the Italian schooling system, compulsory education lasts ten years, from the age of six to sixteen. After five years of primary school, students attend middle – or lower secondary – school for three years, from grade 6 to grade 8.[3] During these years, the curriculum is identical for all students. At the end of the first semester of grade 8, students must formally enroll into the track of upper secondary education they want to pursue. This formal enrollment requires students to complete an application form together with their parents, aided by teachers' non-binding recommendations. Entry into high school is not competitive and does not depend on grades or ability. Thus, the formal application guarantees access to the selected high school. Changing the choice stated in the application is possible, yet time-consuming and infrequent, with no guarantee of acceptance.

3 At the beginning of grade 6, pupils are allocated to classes typically consisting of 18–27 students and stay in the same class for all three years unless they repeat a grade. Classes are formed randomly within schools, ensuring equal distribution of students across classes based on gender, socioeconomic status, and ability level.

Upper secondary education in Italy includes high school, which lasts five years and gives access to tertiary education (namely universities and advanced training institutions), and vocational training, which lasts three or four years and does not give access to tertiary education. Within the high school category, there are three different tracks: lyceum, technical high school, and vocational high school. The lyceum offers six curricula: humanities, scientific, languages, human sciences, artistic, and music and dance. The technical high school provides two curricula: an economic track and a technological track. The vocational high school offers a practical training in two different areas: services (e.g., healthcare, commercial, and agricultural services) and industry and craft (e.g., technical maintenance and assistance).

We divide high schools into two categories: STEM and non-STEM. In the STEM category we include the math-intensive tracks, namely the scientific lyceum and the technological curriculum of the technical high school, which, according to data from the Italian Ministry of Education (MIUR), are strongly correlated with the choice of STEM universities.[4] In fact, the latest MIUR report on the transition from high school to university indicates that 46.9 % of students attending the scientific lyceum and 75.1 % of students attending the technological curriculum of the technical high school enroll in STEM universities. In contrast, only 12.1 % of students attending the human sciences or linguistic lyceum do so.

3. The study design

3.1. Data collection

A total of 548 Italian 8th graders from 27 different classes participated in the study, which was conducted in May 2018. They were recruited from 5 public middle schools in Trento in Northeast Italy. Participation was contingent on obtaining formal permission from the school dean and informed consent from the students' parents. All students invited to participate took part in the study, but we had to exclude 14 students from the analysis due to their severe learning difficulties, as indicated by their teachers. This leaves us with a sample of 534 subjects.

The participating schools were located in well-connected urban areas with easy access to a broad range of high schools, thereby minimizing the potential influence of geographic constraints on students' educational choices. To ensure inclusion and to control for within-school spillovers, all 8th grade classes in the same school participated in the study. Additionally, since the schools were geographically dispersed, we do not need to worry that information about the study spilled over into other schools.

All data were collected in paper-and-pencil format. Data collection was conducted during class time and lasted about 60–70 min per class. Students were informed about the timing of the activity before taking part in it. To ensure privacy, students were physically separated, each seated at her own desk. Prior to data collection, each student received a unique ID number that was taped to her desk and printed on any document she received. Written general instructions were distributed and read aloud by an experimenter. The general instructions emphasized the

[4] The technological curriculum of the technical high school comprises several sub-tracks, some of which are not so closely related to science or technology. We excluded these sub-tracks from the STEM category; specifically, we excluded Graphic Design and Communication; Fash- ion System (not offered in Trento); Agriculture, Agri-Food, and Agroindustry; Transport and Logistics. As a result, our STEM classification includes only technologically and scientifically oriented sub-tracks.

importance of no communication among participants and the anonymity of participation. They also illustrated the structure of the study and the procedure for rewarding.[5]

The variables of interest were collected in different parts. We first assessed each student's objective math ability.[6] Then, we collected incentivized measures of individual math-gender stereotypes and math confidence. Next, we investigated the existence of implicit gender stereotypes. Finally, we administered the survey.[7] Although participants knew from the beginning that there would be different parts, they learned about the content of each part only after having completed the previous one. All parts were conducted consecutively due to restrictions imposed by the school deans, who allowed us to visit each school and all its classes only once.

As we will detail in the following subsection 3.2, students could earn points for their choices. Points were exchanged for euros at the rate of 1 point = €1. Because all participants were minors and could not be paid in cash, their experimental earnings were converted into vouchers to be spent in a Trento bookstore. Only three randomly selected participants per class received the voucher, which was delivered to them approximately one month after completion of the study. Students were informed about this rewarding procedure since the beginning. The random selection was done in each class at the end of the experiment by placing all ID numbers in a box and by asking a volunteer to pick three numbers from the box.

The five schools in our sample display characteristics that are in line with the general population of Italian 8th graders in public schools for the following aspects. First, as we will discuss later, the high school enrollment by gender observed in our sample is comparable to national averages. Second, the proportion of females in our sample (46.4 %) is close to the proportion of Italian female 8th graders (48.1 %; Source: ISTAT). Third, considering that all Italian students conclude their middle school with a nationwide exam, the average scores in this final exam for the five sampled schools are 7.53, 7.65, 7.98, 7.97, and 8.29, where the national average score is 7.68 (Source: MIUR; school year 2017/2018). Finally, the percentages of 8th graders admitted to the final exam in the five schools range between 98.3% and 98.8%, aligning with the national average of 98.3 %.

3.2. Experimental and survey variables: description and procedures

In this section, we describe the variables that we will use in our analysis and the procedures for eliciting them.

3.2.1. Math ability

The first task that student participants performed was a math task, which consisted of ten multiple-choice and open-ended problems adapted from the INVALSI (National Institute for the Evaluation of the Italian Education System) math tests for grade 8.[8] We modified the original INVALSI questions by altering numerical values while preserving their structural integrity and conceptual complexity. This approach allowed us to evaluate the same underlying mathematical competencies as the original INVALSI tests, but minimized any advantage from prior exposure

[5] A translation of the instructions (originally in Italian) can be found in Online Appendix A.

[6] Math ability was measured at the beginning of the study to prevent any potential influence of subsequent tasks on performance.

[7] The experiment included an additional task that will not be analyzed in the present study.

[8] INVALSI tests are standardized assessments administered annually to students attending grades 2, 5, 8, and 10 to evaluate their proficiency in mathematics, Italian, and English.

to specific test questions. We selected INVALSI-style tests to assess students' math ability for several compelling reasons. First, these tests are considered a neutral, objective, and reliable measure of mathematical knowledge.[9] Second, the math performance of girls and boys in the INVALSI tests is not significantly different within the considered macro-geographical area (North-East Italy). Third, students in our sample (none of whom were repeating the 8th grade) receive similar training for the INVALSI tests through exercises drawn from old assessments and specific textbooks. Finally, there is evidence of a positive correlation between teacher-assigned grades and students' INVALSI test scores, which holds true even after dividing the data by gender (Palmerio and Caponera, 2021; Doz, 2023).

Students had 15 min to answer as many problems as they could and thus no competition was involved. They were not allowed to use calculators, but could use the blank space next to each problem to make computations. There were always seven distinct versions of the ten problems to prevent copying. Participants earned 1 point for each problem answered correctly. The performance in this task, namely the number of correctly answered problems, is used as a proxy for participants' objective math ability.

3.2.2. Confidence: beliefs about absolute and relative math ability

High school enrollment and, in particular, the decision to pursue a math-intensive curriculum may depend on the students' beliefs about their own absolute math performance and their relative performance in their peer group. We therefore asked participants (i) to estimate their performance in the math task completed in part 1 and (ii) to rate such a performance as higher than, lower than, or equal to the median performance of others in their class. The exact formulation of this second question, and thus the way in which we conveyed the concept of median performance, was as follows: "Think of your class. Imagine having to rank yourself and your classmates based on the score (number of correct answers provided) obtained in PART 1 quiz, from the lowest to the highest. Now consider the person in the middle position. Do you think that, compared to this person, your score is (please thick the appropriate box): higher/equal/ lower".[10] We gave participants a financial incentive to report beliefs accurately: we paid them 1 point for each belief that turned out to be correct.

3.2.3. Incentivized measure of individual math-gender stereotypes

To obtain an incentivized measure of individual math-gender stereotypes, participants performed a belief elicitation task: they were asked to guess the number of correct answers given, on average, by a female and male student (from a different participating middle school) in the math task. To avoid making gender overly salient, participants were also asked to guess the number of correct answers given by a student in a different class within the same school as well as by a student in a different school. Which gender was presented first was randomized across participants. Herein, we will analyze the beliefs concerning the male and female student's performance since the other beliefs serve the purpose of diverting attention from gender. Each correct guess was rewarded with 1 point.

---

[9] See Doz (2023) as well as INVALSI tests according to INVALSI.

[10] Since we gave students the opportunity to address any questions they had after reading the instructions, we are confident that they understood the concept of median performance.

3.2.4. Implicit math-gender stereotypes

In addition to the incentivized measures of individual stereotypes, we collected unincentivized measures of implicit gender stereotypes using the gender-science Implicit Association Test (IAT) (Greenwald et al., 1998). In the paper-and-pencil (P&P) format administered here, participants received two separate sheets: one stereotype-congruent and the other stereotype-incongruent (the second sheet was handed out only after having collected the first one). Each sheet contained a list of words related to four categories: male, female, science, and liberal arts. Participants were instructed to place each word into the appropriate category by checking the box on the left or the right of the word. In the stereotype-congruent sheet, the categories "male" and "science" were paired on one side of the listed words and the categories "female" and "liberal arts" were paired on the other side. In the stereotype-incongruent sheet, the category pairings were switched. The order of presentation of the two sheets, as well as the left-right location of the pairings, was randomized across participants. For each sheet, participants were given 45 s to categorize as many words as possible starting from the top of the list without skipping any word or correcting mistakes.[11] A measure of implicit math-gender stereotypes is obtained by comparing how correctly participants categorize the words in the two sheets.

We opted for a P&P IAT, a less common alternative to the computerized version, due to logistical constraints at the participating schools. Some schools had only one computer room, which was in continuous use, and the deans were concerned about disrupting the routine activities of other classes. Conducting the sessions in the classrooms allowed us to ensure consistent test administration and to make the most efficient use of the available time by avoiding transitions to the computer room. While the computerized IAT measures reaction time precisely, the P&P version approximates response speed by considering the number of items correctly categorized within a given time. Several studies have confirmed that the P&P IAT yields results that significantly correlate with those of the computerized version (Lemm et al., 2008; Bardin et al., 2016; Filleul et al., 2023), making it a suitable and reliable alternative for our study.

3.2.5. Survey measures of explicit math-gender stereotypes

The survey included three questions asking participants who is better at math between girls and boys according to (i) themselves, (ii) their teachers, and (iii) their classmates. These questions are commonly used in the psychological literature to assess the explicit endorsement of gender stereotypes about math (Passolunghi et al., 2014). Answers to all three questions were on a 5-point Likert scale from "girls are definitely better" to "boys are definitely better".

3.2.6. Reported high school track and other math-related survey measures

In the survey, participants were asked to report the track (lyceum, technical, or vocational) and the name of the high school they enrolled in, the specific curriculum they chose, as well as the official, albeit non-binding, recommendation given by their teachers. We also collected three math-related measures in the survey: the participants' degree of confidence in their math skills, their perceived value of math, and the interest in math conveyed by the school. Students' math

[11] Participants were allowed to familiarize with the task by classifying words related to the categories flowers, insects, positive adjectives, and negative adjectives. The test was administered after having verified that all participants correctly understood the task.

confidence was assessed with the aid of the scale used in the Trends in International Mathematics and Science Study (TIMSS, 2015, 2019, grade 8). This scale includes nine statements with a 5-point Likert response format ranging from 1 (strongly disagree) to 5 (strongly agree). The perceived value of math was measured by Lim and Chapman (2013)'s subscale comprising five statements to which participants responded on a 5-point Likert scale.[12] Finally, the perceived contribution of math courses to the development of interest in math was assessed by asking students to indicate, on a 4-point Likert scale, how much they thought the courses helped them be aware of the applications of math in real life, realize how math is relevant for everyday decisions, get excited about math.

4. Results

Table 1 shows the composition of our sample in terms of gender for each school and overall. Out of 534 students, 248 are females and 286 are males.

4.1. Reported high school track enrollment

The percentages of female and male students who reported enrolling in a STEM high school are displayed in Fig. 1. Nearly half of the boys reported STEM enrollment, while only 28.2 % of girls did so. A test of proportions indicates that boys are significantly more likely than girls to report STEM enrollment (Pearson χ^2 test, p-value < 0.001).[13] The gender difference in high school STEM enrollment observed in our sample mirrors that at the national level: MIUR data for the school year 2017–2018 show that 54.13 % of boys and 25.33 % of girls were enrolled in a STEM high school.[14] Before investigating what can help explain this lopsided choice, we provide a descriptive analysis of our experimental and survey variables and test for possible gender differences in them.

4.2. Objective and self-perceived math ability

We begin the analysis by examining female and male students' actual and expected performance in the math task.

4.2.1. Objective math ability

Fig. 2 presents the distribution of correct answers in the math task for boys and girls separately (panel a) as well as the corresponding descriptive statistics (panel b). The mean number of math problems answered correctly is 4.654 for male students and 4.480 for female students, and median values equal 4.000 for both genders.

The answers of boys are slightly more dispersed than those of girls, as evidenced by the standard deviation. Regarding symmetry, both distributions are positively skewed, with a mass shift to the left. In accordance with previous studies, the distributions of correct answers of boys and girls do not statistically differ, neither in their central tendency (Wilcoxon Rank Sum test,

[12] See the instructions in Online Appendix A for the statements composing the TIMSS scale and the Lim and Chapman's subscale.

[13] All statistical tests presented in the paper are two-sided.

[14] See the Italian Ministry of Education data portal: Student Registry Area.

Table 1
Gender composition of our sample.

School	No. of Classes	Females	Males	Tot
A	5	58 (53.7 %)	50 (46.3 %)	108
B	6	62 (47.0 %)	70 (53.0 %)	132
C	6	49 (48.5 %)	52 (51.5 %)	101
D	5	38 (40.4 %)	56 (59.6 %)	94
E	5	41 (41.4 %)	58 (58.6 %)	99
TOT	27	248 (46.4 %)	286 (53.6 %)	534

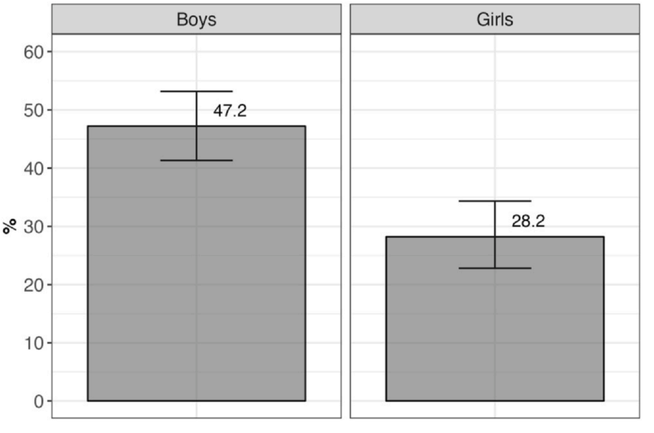


Fig. 1. Frequency of enrollment in STEM high school by gender.
Notes: The height of the bars and the number above the bars indicate the percentage of boys and girls who reported enrolling in a STEM high school. The figure also shows confidence intervals from the Pearson χ^2 test.

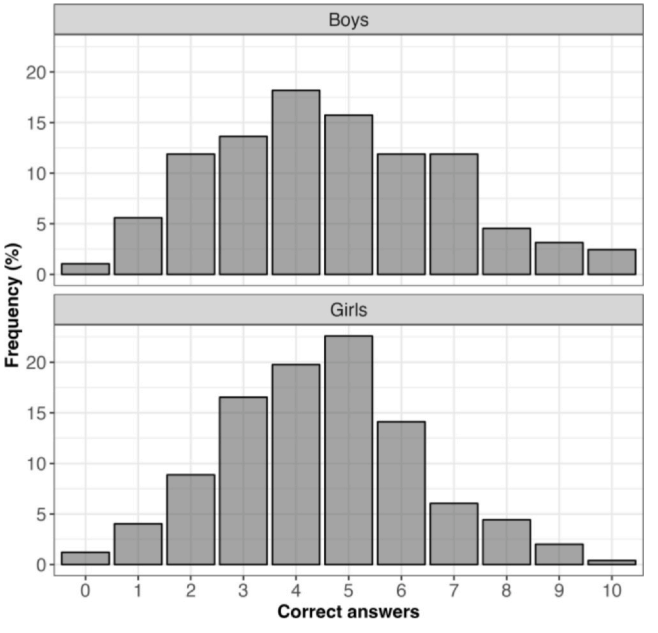
henceforth WRS, p-value=0.526) nor in the shape of their cumulative distribution function (Kolmogorov-Smirnov test, p-value=0.219). Thus, overall, males and females do not differ significantly in objective math ability.

Table 2 displays the distribution of boys and girls across quartiles of math performance, which will serve as a proxy for math ability in our regression analysis. ‘MATH Q1’ corresponds to the lowest quartile and ’MATH Q4’ to the highest quartile. As the table shows, boys are more likely to be in the extreme quartiles of math performance compared to girls, confirming that boys’ math ability exhibits higher variance. Notably, there is a sharp difference between genders in the top quartile, with 22.03 % of boys compared to 12.90 % of girls.

4.2.2. *Beliefs about absolute and relative math ability*

Fig. 3 depicts the distribution of the participants’ beliefs about their own math performance (panel a), separated by gender, displaying descriptive statistics in tabular form (panel b). The median equals 7 for both genders, but the mean estimate is slightly higher for male students. The figure makes it apparent that boys’ beliefs are more concentrated on the high values of the

(a) Distribution of correct answers in the math task



(b) Descriptive statistics of correct answers

Gender	N	Mean	Median	SD	Skewness
Boys	286	4.654	4.000	2.230	0.289
Girls	248	4.480	4.000	1.885	0.191

Fig. 2. Math performance by gender.

distribution than girls' beliefs, as testified also by the higher negative skewness. The two cumulative distributions differ in shape (Kolmogorov-Smirnov test, p-value=0.003) and central tendency (WRS test, p-value < 0.001).[15]

Turning to the second measure of confidence that we elicited, i.e., the perceived position relative to the median, Table 3 presents the percentages of male and female students who believe to have performed in the math task worse than, the same as, and better than the median classmate. There is a significant gender difference in expected relative math performance

[15] The correctness of these beliefs can be used to identify potential overestimation of one's actual math ability, which Moore and Healy (2008) consider as a variety of overconfidence. Wilcoxon Signed Rank (WSR) tests show that, for both genders, this type of overconfidence is significantly different from zero (both p-values < 0.001). However, the level of overestimation is marginally significantly higher for boys than for girls (WRS test, p-value=0.071).

Table 2
Distribution of boys and girls by quartiles of math performance.

Gender	MATH Q1	MATH Q2	MATH Q3	MATH Q4
Boys	32.17 %	18.18 %	27.62 %	22.03 %
Girls	30.65 %	19.76 %	36.69 %	12.90 %

(Pearson χ^2 test, p-value < 0.001): most boys believe they are better than the median, and most girls believe they are as good as the median.[16]

4.3. Individual math-gender stereotypes

We now turn our attention to the participants' individual math-gender stereotypes as measured by the incentivized task, the IAT, and the survey questions.

4.3.1. Incentivized measure of individual math-gender stereotypes

We obtained an incentivized measure of individual math-gender stereotypes via a belief elicitation task. Fig. 4 displays the distribution of the individual-level difference between expected correct answers given by a male and those given by a female, separately for each gender (panel a); the corresponding descriptive statistics are given in panel (b).

For both genders, the expected mean difference in math performance between male and female students is negative and significantly different from zero (WSR test, p-values < 0.001 for both genders), indicating that, on average, participants expect female students to perform better than male students. This suggests that neither gender adheres to traditional math-gender stereotypes. Rather, both boys and girls exhibit a reverse stereotype, although the magnitude of the expected difference is modest. These expectations contrast with actual math performance observed in our sample, where boys slightly outperform girls, with a mean difference of 0.174 (see Fig. 2b). Moreover, belief-based performance expectations do not significantly differ by participant gender (WRS test, p-value=0.687).

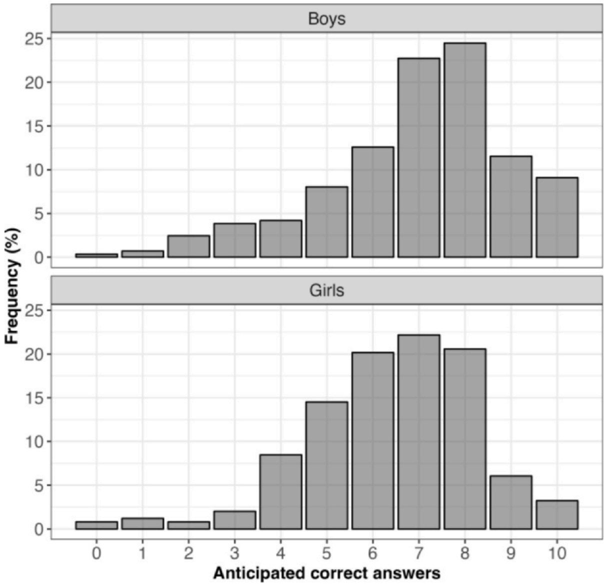
4.3.2. Individual math-gender stereotypes at the implicit level

To measure implicit math-gender stereotypes in our sample, we compute for each participant a (σ) IAT score following the "product: square root of difference" procedure recommended by Lemm et al. (2008).[17] Larger scores denote larger implicit math-gender stereotypes, i.e., more

[16] To assess the accuracy of these beliefs, we compare the students' expected relative performance with their actual relative performance. This allows us to check for potential over-placement, which Moore and Healy (2008) identify as a further measure of overconfidence. We find that 41.1 % (46.6 %) of girls (boys) are accurate in their evaluations. However, more boys than girls overestimate their relative performance (37.8 % vs 33.8 %) and more girls than boys underestimate it (25.1 % vs 15.6 %). The gender difference in under-placement is statistically significant (Pearson χ^2 test, p-value=0.033).

[17] Lemm et al. (2008) compare the performance of alternative scoring procedures and conclude that the 'product: square root of difference' algorithm outperforms other measures, overall. To compute the IAT score, we proceed as follows. First, we take A as the number of correct categorizations in the stereotype-congruent condition and B as the number of correct categorizations in the stereotype-incongruent condition. Then, we define $X = max\{A, B\}$ and $Y=min\{A, B\}$. The IAT score is calculated as $\sigma = X/Y \times \sqrt{X - Y}$. To retain the directionality of the bias (females are associated with science less than males), the resulting score is multiplied by –1 if $B > A$ (i.e., if a participant scores higher on the sheet that pairs 'female' with 'science'). To improve the robustness of our results, we consider only subjects with at least 20 % correct categorizations. However, results do not change when the entire sample is considered.

(a) Distribution of expected correct answers in the math task



(b) Descriptive statistics of expected correct answers

Gender	N	Mean	Median	SD	Skewness
Boys	286	6.993	7.000	2.003	-0.803
Girls	248	6.427	7.000	1.823	-0.647

Fig. 3. Beliefs on own math performance by gender.

automatic associations between male gender and math. We find that mean IAT scores are 1.474 for girls and 0.172 for boys. According to WSR tests, the girls' score is significantly different from zero (p-value < 0.001), whereas the boys' score is not (p-value=0.680). A WRS test indicates that there is a significant gender difference in IAT scores (p-value < 0.001). Thus, overall, girls (but not boys) display significant implicit math-gender stereotypes.

4.3.3. Individual math-gender stereotypes at the explicit level

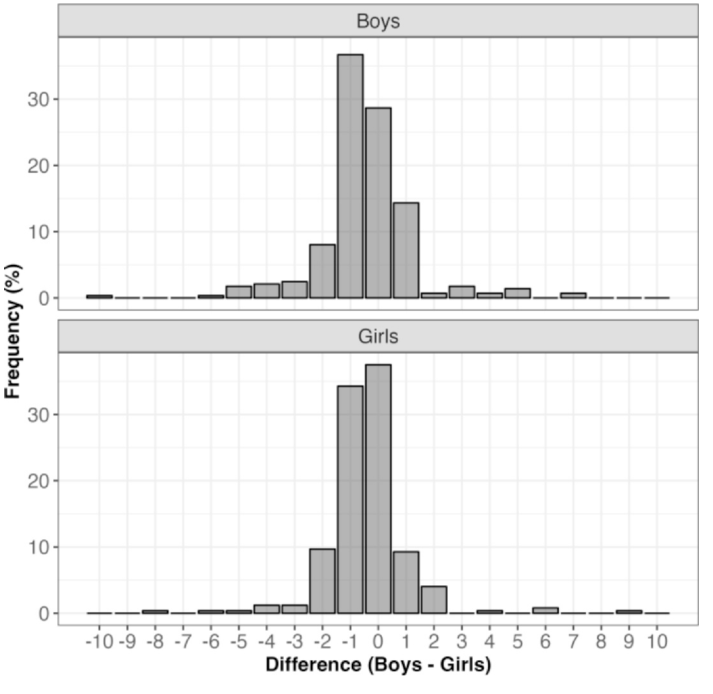
To evaluate participants' explicit endorsement of math-gender stereotypes, we collected responses to three survey questions that assessed (i) participants' own beliefs about whether males or females are better at math, (ii) their perceptions of their teachers' beliefs, and (iii) their perceptions of their classmates' beliefs. Responses were recorded on a scale ranging from −2 ("girls are definitely better") to +2 ("boys are definitely better"), with 0 indicating no perceived bias.

The results show a clear pattern of gendered bias. On average, male students believe that boys are better at math (mean=0.331), whereas female students believe that girls are better (mean=−0.071). A similar pattern emerges in perceptions of classmates' beliefs (means: 0.257

Table 3
Own expected relative performance in the math task by gender.

Gender	Worse	Same	Better
Boys	13.0 %	36.6 %	50.4 %
Girls	14.7 %	58.4 %	26.8 %

(a) Distribution of the expected difference in correct answers between boys and girls



(b) Descriptive statistics of the expected difference

Gender	N	Mean	Median	SD	Skewness
Boys	286	-0.472	-1.000	1.754	0.049
Girls	248	-0.423	0.000	1.517	0.774

Fig. 4. Beliefs on the difference in math performance between males and females by gender.

for boys, −0.179 for girls). In contrast, both genders perceive teachers as slightly biased in favor of girls, though the magnitude is modest (−0.118 for boys and −0.108 for girls).

These questions were designed to capture explicit gender-related messages to which students are exposed in their school environment. The responses exhibit high internal consistency (Cronbach's alpha = 0.860), allowing us to construct an aggregate measure. Specifically, we average the three responses to compute an *explicit bias index*, which serves as our main measure of explicit stereotype endorsement. The mean index score is −0.119 for girls and + 0.157 for boys; both values are significantly different from zero (WSR test, p-value=0.034 for girls and 0.016 for boys), These results indicate that both genders exhibit a moderate but statistically significant explicit bias in favor of their own gender.

4.4. Relationship between math ability, individual math-gender stereotypes, and reported high school enrollment

This section presents our main findings on the relationship between our variables of interest and reported high school enrollment. To this end, we estimate the following general model:

$$Y_{ic} = \beta_0 + \gamma_1 \text{MATH}_{ic} + \gamma_2 \text{CONF}_{ic} + \gamma_3 \text{MGS}_{ic} + \delta_c + \mu_{ic}$$

where *i* indicates the student and *c* her/his class. The dependent variable Y_{ic} is a binary variable that equals 1 if student *i* in class *c* reported STEM enrollment and 0 if the student reported non-STEM enrollment. The set of explanatory variables include: *i*'s objective math ability (MATH_{ic}), *i*'s confidence or self-perceived math ability (CONF_{ic}), and *i*'s math-gender stereotypes (MGS_{ic});[18] δ_c captures class fixed effects and μ_{ic} is the error term.[19]

Table 4 reports the outcomes of linear regressions (OLS) estimating the coefficients of our model separately for boys and girls.[20] In our specifications, we first consider each of the main explanatory variables in isolation, controlling for math ability, and then we assess whether the results hold when the effects of all the explanatory variables are estimated simultaneously. Model (1) focuses solely on students' math ability (MATH), divided into math performance quartiles as shown in Table 2. We opted for this specification rather than a continuous one to better capture potential non-linear effects of math skills and to allow for a direct comparison between students with math skills above and below the median. We assume this variable is predetermined and unaffected by the track a student enrolls in because the measurement of math ability occurred close to the enrollment decision and improving math skills requires a substantial investment of time and effort. Model (2) adds measures of students' self-perceived math ability (CONF) to Model (1). Specifically, it includes the variable 'MATH belief', representing the student's predicted number of correct answers in the math test, and 'MATH median', a dichotomous variable equal to 1 if the student believes to be strictly above the class median in the math test. As the correlation matrix in Online Appendix B shows, these variables are positively related to objective math ability. Model (3) extends Model (1) by incorporating measures of individual math-gender stereotypes (MGS). 'Diff Exp Performance' denotes the

[18] Table B.1 in Online Appendix B presents the summary statistics of the variables used in the regressions.

[19] In the regressions, we control for class fixed effects, but the results remain substantively unchanged when these fixed effects are omitted.

[20] Table B.2 in Online Appendix B presents the coefficient estimates from a generalized linear model (Logit) and shows that results are robust to this different model specification.

Table 4
Determinants of STEM enrollment separately for boys and girls.

Boys				
	(1)	(2)	(3)	(4)
Intercept	0.408 (0.131)**	0.151 (0.175)	0.445 (0.130)***	0.141 (0.182)
MATH Q2	0.155 (0.082)°	0.103 (0.085)	0.119 (0.081)	0.055 (0.082)
MATH Q3	0.332 (0.077)***	0.279 (0.087)**	0.300 (0.082)***	0.223 (0.091)*
MATH Q4	0.564 (0.070)***	0.457 (0.090)***	0.563 (0.074)***	0.443 (0.092)***
MATH belief		0.035 (0.017)*		0.045 (0.018)*
MATH median		0.060 (0.064)		0.030 (0.064)
Diff Exp Performance			−0.001 (0.015)	−0.007 (0.016)
Own implicit bias			0.014 (0.006)*	0.013 (0.006)*
Explicit bias index			0.073 (0.034)*	0.088 (0.034)*
N	286	262	261	239
Adj R^2	0.254	0.277	0.285	0.319
F-stat (p-value)	< 0.001	< 0.001	< 0.001	< 0.001
Girls				
	(1)	(2)	(3)	(4)
Intercept	0.101 (0.129)	0.000 (0.151)	0.161 (0.131)	0.074 (0.166)
MATH Q2	0.129 (0.082)	0.174 (0.080)*	0.110 (0.089)	0.146 (0.085)°
MATH Q3	0.162 (0.073)*	0.161 (0.072)*	0.140 (0.081)°	0.134 (0.079)°
MATH Q4	0.168 (0.104)	0.106 (0.116)	0.143 (0.110)	0.085 (0.122)
MATH belief		0.012 (0.016)		0.008 (0.018)
MATH median		0.176 (0.073)*		0.195 (0.078)*
Diff Exp Performance			−0.039 (0.018)*	−0.034 (0.017)*
Own implicit bias			−0.007 (0.007)	−0.006 (0.007)
Explicit bias index			0.018 (0.042)	−0.006 (0.042)
N	248	231	234	218
Adj R^2	0.156	0.219	0.143	0.201
F-stat (p-value)	< 0.001	< 0.001	< 0.001	< 0.001

Notes: Coefficient estimates are from linear regressions performed separately by gender. The dependent variable equals 1 if the student reported enrolling in a STEM track and 0 otherwise. The regressions include class fixed effects that are not shown. Robust standard errors are reported in parentheses. Symbols ***, **, *, and ° denote statistical significance at the 0.1 %, 1 %, 5 %, and 10 %, respectively.

participants' expected difference in math performance between male and female students. 'Own implicit bias' refers to the participants' IAT scores. 'Explicit bias index' is the index of explicit math-gender stereotypes as defined in Section 4.3. Model (4) is the full model, presenting joint estimates of all explanatory variables.[21]

[2]In all models, the estimated coefficients for math ability indicate that top-performing boys are more likely to report enrollment in a STEM high school than boys in the baseline quartile, while this does not hold true for top-performing girls. Linear restriction tests on the estimated coefficients further show that boys in the highest quartile are significantly more likely to report STEM enrollment than those in the other two quartiles across all specifications (χ^2 test, all p-values ≤ 0.026). Conversely, girls in the highest quartile do not exhibit a significantly higher likelihood of reporting STEM enrollment compared to those in the other quartiles across all specifications (χ^2 test, all p-values ≥0.577). We summarize these findings in our first result.

[21] We address potential multicollinearity among the regressors in Online Appendix B.

Result 1. *Objectively higher math abilities do not significantly increase the likelihood of STEM enrollment for girls, whereas they strongly predict increased STEM participation among boys.*

When we consider self-perceived math ability, Models (2) and (4) show a distinct pattern by gender: the coefficient of 'MATH belief' is positive and significant for boys but not for girls, while the coefficient of 'MATH median' is positive and significant for girls but not for boys. This brings us to our second result.

Result 2. *Controlling for math ability, girls are significantly more likely to report STEM enrollment the better at math they perceive themselves in relative terms, while what matters the most for boys is their self-perceived math ability in absolute terms.*

Turning to the three considered measures of individual stereotypes, Models (3) and (4) indicate that only the coefficient of 'Diff Exp Performance', representing the difference between the expected performance of boys and girls in the math task, is negative and significant at the 5 % level for girls. 'Own implicit bias' and 'Explicit bias index' positively and significantly affect boys' STEM enrollment. This leads to our third result.

Result 3. *Controlling for math ability, girls holding more gender-stereotyped incentivized beliefs about math ability are significantly less likely to report STEM enrollment. For boys, stronger implicit and explicit math-gender stereotypes are significantly associated with a higher likelihood of reporting STEM enrollment.*

4.5. Survey measures related to math

In this section, we focus on the three math-related questions included in the survey. The survey-based measure of math confidence assessed how confident students felt about their competence in math in terms of their level of agreement (on a 5-point scale) with nine statements about math. For each student, we compute a score by averaging the nine answers. Averaging them over students, we obtain a mean confidence score of 3.044 for girls and 3.330 for boys.[22] Both girls and boys display an overall moderate confidence in math, but girls report significantly lower confidence than boys (WRS test, p-value=0.003).

The survey also asked for students' level of agreement with five statements regarding the worth of math to their lives. Once again, individual scores are obtained by averaging the 5-point Likert scores on the individual questions and mean scores by averaging the individual scores. Both girls and boys perceive math as valuable, with mean scores respectively of 3.816 and 3.902, and there is no gender difference (WRS test, p-value=0.183).

Finally, the survey included three questions (with responses on a 4-point scale) to measure the extent to which the math courses attended at school helped students develop an interest in math. The three answers are combined into a single individual average score. The overall mean score is 2.606 for girls and 2.727 for boys. Hence, both genders perceive a positive impact of math courses on math interest, but girls to a significantly lower extent than boys (WRS test, p-value=0.036).

To summarize, the math-related (non-incentivized) survey measures confirm our previous results based on incentivized measures of confidence and additionally indicate that both genders

[22] Statements expressing negative sentiments are reverse-coded.

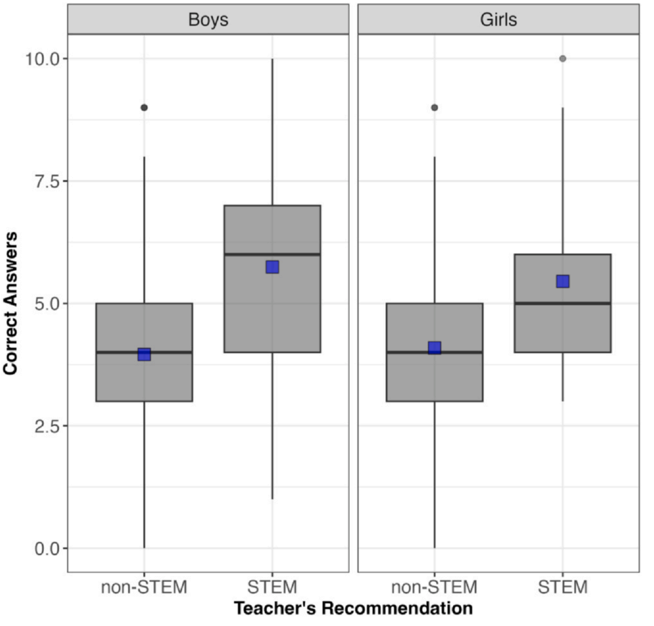


Fig. 5. Math performance by gender and track recommended by teachers.

perceive math as valuable and school as important to the development of their interest in math, although girls report a significantly lower score on the questions related to school. When comparing students by reported high school track, we observe that compared to their counterparts who reported non-STEM enrollment, girls and boys who opted for STEM enrollment (i) are significantly more confident in math, (ii) regard math as significantly more valuable, and (iii) perceive math courses as a greater boost to their math interest (WRS test, p-value < 0.001 for all pairwise comparisons).

4.6. *Teachers' recommendations*

Teachers play a crucial role in shaping students' educational trajectories, particularly at critical junctures such as the transition to high school. In this section, we examine whether gender bias is present in teachers' recommendations for math-intensive tracks, controlling for students' objective math abilities. Teachers' high school track recommendations strongly correlate with students' reported enrollment, even though the recommendations are not binding. Specifically, 87.5 % of girls and 89.4 % of boys who were recommended for a STEM track report following the recommendation. Among those not recommended for STEM, only 14.2 % of girls and 22.0 % of boys report not following the recommendation. The difference in the likelihood of STEM enrollment between those recommended and those not recommended is statistically significant for both genders (both χ^2 test p-values < 0.001).

Controlling for students' math ability, Fig. 5 depicts the distribution of correct answers in the math task for both boys and girls, grouped by the track (non-STEM vs STEM) recommended by teachers. Teachers tend to recommend a STEM track to students with higher math performance, regardless of gender. WRS tests confirm that the difference in math performance between students recommended for a STEM track and those recommended for a non-STEM track is statistically significant for both boys and girls (both p-values < 0.001). We can thus state that the teachers' track recommendations differentiate based not on gender, but rather on objective math ability, as assessed by the modified INVALSI tests. This result (i) validates the use of our tests as a reliable measure of math ability and (ii) rules out gender bias in teachers' guidance as a contributing factor to the underrepresentation of girls in STEM fields within our study context.

5. Discussion and concluding remarks

The reason why girls do not choose the most rewarding STEM educational tracks in terms of future wages and labor market opportunities is still debatable. In this paper, we aimed at determining how a number of gendered math-related factors, collected through a lab-in-the-field experiment, are related to the high school track in which Italian 8th graders report to have enrolled. By addressing this research question, we contribute to a small but growing literature that links individual characteristics elicited in a controlled experimental setting to real-life decisions.

Math ability and skills are deemed necessary for pursuing math-intensive studies, and our experiment shows, in line with previous literature, that boys and girls perform equally well on math tests. Despite this equality, girls in our sample are significantly less likely to report enrollment in a math-intensive high school track.

The main result of our regression analysis is that, conditional on math performance, girls are significantly more likely to report STEM enrollment, the more they think they are better than the median classmate in terms of math ability and the stronger are their counter-stereotypical incentivized beliefs that females on average outperform males in the math task. Higher objective math ability (measured by the number of correct answers in the math task) and higher self-perceived math ability in absolute terms (namely higher expected number of correct answers in the math task) do not significantly impact girls' STEM enrollment, whereas they are significant predictors—along with implicit and explicit math-gender stereotypes—of boys' STEM enrollment. The observation that beliefs about relative math performance are relevant for girls (but not boys) and beliefs about absolute math performance are relevant for boys (but not girls) is consistent with psychological theories suggesting that girls and boys frame their math ability differently and focus on distinct aspects when making educational choices (Eccles, 1994; Loyalka et al., 2017).

In addition to the key findings from our regression analysis, we document other notable gender differences. First, compared to boys, girls tend to have lower confidence in their math ability. Second, girls, but not boys, display math-gender stereotypical associations at the implicit level. Conversely, both genders exhibit a strong ingroup bias at the explicit level. The result that girls hold significant implicit math-gender stereotypes, even in the absence of explicit traditional stereotypes, aligns with previous evidence on Italian students (Galdi et al., 2014; Passolunghi et al., 2014) and supports the hypothesis that implicit stereotypes may develop earlier and independently of explicit ones (Passolunghi et al., 2014; Vuletich et al., 2020).

While existing research has documented that teachers' gender stereotypes can shape girls' educational trajectories by discouraging math-intensive track selection (Carlana, 2019), we find no such evidence in our setting. In our sample, teachers recommend STEM tracks to the students with better math performance, without differentiating on the basis of gender.

Overall, our study provides evidence that perceived comparisons with peers and gender counter-stereotypical beliefs about math are likely to affect girls' educational track decisions.

As a consequence, any policy aimed at improving the relative expectations of girls may increase the number of female students who enroll in a math-intensive high school track. The systematic disclosure of the girls' actual good math performance compared to boys may for instance serve this purpose. Another option (often brought up in the literature and recently investigated by Breda et al., 2023) is to provide girls with information highlighting female role models in math and science. Such role models may indeed be seen as an affirmation by young girls in school (Mouganie and Wang, 2020), namely as an indication that they can perform equally well and, by this means, role models may boost girls' expectations about their relative math ability. Obviously, encouraging more girls into math-intensive high schools may not be a sufficient remedy for reducing the gender gap in STEM majors and careers because, e.g., females should then have the same chances to succeed as males and not be discriminated against in the job market. Yet, in educational systems characterized by early tracking, understanding what causes math-talented girls to self-select out of math-intensive high schools and identifying appropriate interventions to address these causes should help narrow the STEM gender gap at the outset and thus offer young women the possibility of not lagging behind in math and science subjects.

Funding

This work was supported by the grant RB2017LEVATI, sponsored by the University of Verona.

Acknowledgments

We are grateful to participants at the Second Meeting of the Behavioral and Experimental Economics Network, the 2019 VELE Workshop on Lab and Field Experiments, the 2022 European ESA Meeting, and the University of Trento seminar series for their helpful comments and suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jpolmod.2026.107046.

References

Alan, S., Ertac, S., & Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics, 100*, 876–890.

Bardin, B., Perrissol, S., Py, J., Fos, Y., & Souchon, N. (2016). Testing of a paper- and-pencil personalized single category implicit association test (SC-IAT-P). *International Review of Social Psychology, 29*, 31–44.

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

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review, 109*, 739–773.

Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2023). How effective are female role models in steering girls towards STEM? Evidence from French high schools. *The Economic Journal, 133*, 1773–1809.

Breda, T., & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science, 353*, 474–478.

Breda, T., & Ly, S. T. (2015). Professors in core science are not always biased against women: Evidence from France. *American Economic Journal: Applied Economics, 7*, 53–75.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics, 129*, 1409–1447.

Buser, T., Peter, N., & Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review: Papers & Proceedings, 107*, 125–130.

Buser, T., van den Assem, M. J., & van Dolder, D. (2023). Gender and willingness to compete for high stakes. *Journal of Economic Behavior and Organization, 206*, 350–370.

Carpenter, J., Frank, R., & Huet-Vaughn, E. (2018). Gender differences in interpersonal and intrapersonal competitive behavior. *Journal of Behavioral and Experimental Economics, 77*, 170–176.

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

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

Contini, D., Di Tommaso, M. L., & Mendolia, S. (2017). The gender gap in mathematics achievement: Evidence from Italian data. *Economics of Education Review, 58*, 32–42.

Dahlbom, L., Jakobsson, A., Jakobsson, N., & Kotsadam, A. (2011). Gender and overconfidence: Are girls really overconfident? *Applied Economics Letters, 18*, 325–327.

Doz, D. (2023). Factors influencing teachers' grading standards in mathematics. *Oxford Review of Education, 49*, 819–837.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly, 18*, 585–609.

EIGE – European Institute for Gender Equality (2017). *Economic benefits of gender equality in the EU: how gender equality in STEM education leads to economic growth.* Publications Office of the European Union.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 103–127.

Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *Quarterly Journal of Economics, 137*, 1345–1381.

Filleul, V., d'Arripe-Longueville, F., Meinadier, E., Maillot, J., Chan, D. K.-C., Scoffier-Mériaux, S., & Corrion, K. (2023). Development of a French paper-and-pencil implicit association test to measure athletes' implicit doping attitude (IAT-Dop). *International Review of Social Psychology, 36*, 1–19.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*, 25–44.

Francesconi, M., & Parey, M. (2018). Early gender gaps among university graduates. *European Economic Review, 109*, 63–82.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2*, 210–240.

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development, 85*, 250–263.

Granato, S. (2023). Early influences and the choice of college major: Can policies reduce the gender gap in scientific curricula (STEM)? *Journal of Policy Modeling, 45*, 494–521.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494–495.

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

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

Lemm, K., Lane, K., Sattler, D., Khan, S., & Nosek, B. (2008). Assessing implicit cognitions with a paper-format implicit association test. In M. Morrison, & T. Morrison (Eds.). *The Psychology of Modern Prejudice* (pp. 123–146). Nova Science Publishers.

Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics, 82*, 145–164.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135.

Loyalka, P.K., Maani, M., Qu, Y., Sylvia, S., 2017. Absolute versus comparative advantage: consequences for gender gaps in STEM and college access. SSRN Working Paper 2908533.

Matteucci, M., & Mignani, S. (2021). Investigating gender differences in mathematics by performance levels in the Italian school system. *Studies in Educational Evaluation, 70*, Article 101022.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*, 502–517.

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

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies, 87*, 2777–2826.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*, 747–759.

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives, 24*, 129–144.

OECD (2015). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence.* PISA: OECD Publishing.

Palmerio, L., & Caponera, E. (2021). TIMSS Mathematics achievement, school grades and national test scores: different or similar measures of student learning? In P. Falzetti (Ed.). *INVALSI data: methodologies and results* (pp. 37–56). Franco Angeli.

Paryavi, M., Campos, F., & Santos, I. (2025). Women don't avoid competition, they avoid competing against men: Experimental evidence from Kenya. *Journal of Behavioral and Experimental Economics, 116*, Article 102353.

Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math-gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70–76.

Plante, I., O'Keefe, P. A., Aronson, J., Fr´echette-Simard, C., & Goulet, M. (2019). The interest gap: How gender stereotype endorsement about abilities predicts differences in academic interests. *Social Psychology of Education, 22*, 227–245.

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

Rapoport, B., & Thibout, C. (2018). Why do boys and girls make different educational choices? The influence of expected earnings and test scores. *Economics of Education Review, 62*, 205–229.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, 111*, Article 44034408.

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

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

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102*, 947–963.

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*, 581–593.

Vuletich, H. A., Kurtz-Costes, B., Cooley, E., & Payne, B. K. (2020). Math and language gender stereotypes: Age and gender differences in implicit biases and explicit beliefs. *PLoS One, 15*, Article e0238230.