Working Paper Series
Department of Economics
University of Verona

Math ability, gender stereotypes about math ability, and educational choices. Combining experimental and survey data

Dominique Cappelletti, Maria Vittoria Levati, Matteo Ploner

WP Number: 7 June 2022

ISSN: 2036-2919 (paper), 2036-4679 (online)

Math ability, gender stereotypes about math ability, and educational choices.

Combining experimental and survey data

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

[a] *Department of Economics, University of Verona, Italy*

[b] *FBK-IRVAPP, Trento, Italy*

[c] *Cognitive and Experimental Economics Laboratory (CEEL), University of Trento, Italy*

Abstract

The underrepresentation of females in STEM fields negatively affects productivity growth and contributes to labour market inequalities. In countries where children are tracked in 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 comparisons with peers are predictors of the likelihood that girls choose a math-intensive track during high school. Policy initiatives improving girls' expectations about their relative math performance may thus encourage female students to pursue a STEM track.

JEL Classification: C93, J16, J24, I24

Keywords: School choice; Math ability; Gender stereotypes; Beliefs; STEM

We gratefully acknowledge the financial support from the grant "Ricerca di base", RB2017LEVATI, sponsored by the University of Verona. We would like to thank Ivan Soraperra for productive discussions related to this project and Valeria Fanghella for her invaluable assistance during data collection. The paper has benefited from comments by participants at the second meeting of the Behavioral and Experimental Economics Network, the 2019 VELE Workshop on Lab and Field Experiments, and the University of Trento seminar series. Contact the corresponding author, M. Vittoria Levati, at vittoria.levati@univr.it.

1 Introduction

Skills in science, technology, engineering, and mathematics (STEM) are becoming increasingly important to society as economies compete in a global market requiring constant innovation and technological discoveries. In most developed countries, the demand for STEM-educated labour grows at a faster-than-average rate (Caprile et al., 2015) and the COVID-19 pandemic has accelerated the need for digital competence. Employers seek workers with STEM skills and knowledge not only to drive innovation, but also to increase productivity in all fields. Despite the promising job prospects, the increasing demand for STEM skills is not matched by an analogous increase in the number of youngsters pursuing studies in these fields. The problem is particularly pronounced for female students, who are much less likely than males to enrol in STEM tracks. In the European Union, for example, females account for only 25.9% of tertiary education students in engineering, manufacturing and construction and for 38.9% of tertiary students in natural sciences, mathematics and statistics.[1]

These marked gender differences in educational choices have not only detrimental effects on aggregate productivity,[2] but also significant implications for labour market inequality. Kirkeboen et al. (2016), for instance, estimate that students majoring in science rather than humanities almost triple their early career earnings in Norway and, importantly, several studies provide evidence that a substantial part of the gender wage gap is accounted for by gender differences in college and university majors (see, e.g., McDonald and Thornton, 2007, and Blau and Kahn, 2017 for the US; Francesconi and Parey, 2018 for Germany; Card and Payne, 2021 for Canada). It is therefore not surprising that the underrepresentation of women in STEM education has become a major concern for policy-makers and other stakeholders, who call for ways to reduce girls' barriers to enter math- and science-intensive studies. This interest is mirrored in the academic literature, with much effort devoted to understanding the drivers of gendered educational pathways. Reliably identifying such drivers by relating experimentally measured objective and subjective characteristics of male and female students

[1] See Eurostat Tertiary education statistics at `https://ec.europa.eu/eurostat/statistics-explained/index.php/Tertiary_education_statistics`.

[2] Reducing the gender gap would foster economic growth mainly by increasing the labour supply in high-added value sectors. It has been estimated that in the European Union this would result in an increase in the per capita GDP by 0.7-0.9% in one decade and by 2.2-3.0% in three decades (EIGE, 2017).

to their actual educational choices is a critical yet insufficiently studied area.

In the present paper, we make a step in this direction: to assess the relevance of a number of gendered math-related factors (namely ability, confidence, stereotypes) for educational choices of Italian 8th graders, we combine a lab-in-the-field experiment designed to measure these factors with an in-depth survey asking the students about their actual high (or upper secondary) school track choice, among other pieces of information. A key feature of the Italian educational system is its early tracking, with specialised curricula that vary considerably in their math and science intensity already in the high school, and Italian students make the high school track choice during grade 8 (the last year of middle school, our sample of analysis), usually at age 14. This choice is hard to reverse and greatly affects tertiary educational choices and, in turn, labour market opportunities. Data from the Italian Ministry of Education show indeed that most of the students attending math- and science-intensive high school curricula enrol in STEM universities, and a recent study by Granato (2020) indicates that about half of the gender gap in STEM graduation in Italy is attributable to gendered high school choices. Thus, in educational systems characterised by early tracking (like the Italian one), if one wants to help shape policy interventions aimed at encouraging girls' participation in STEM studies, one ought to understand what contributes to steer girls away from math- and science-intensive fields before educational segregation starts.

Gender differences in math ability are amongst the factors suggested to influence the type of pursued education. Girls are found to underperform boys in math standardised tests at all school grades, with the gap widening as children get older (for evidence on the US and Europe see, respectively, Fryer and Levitt, 2010 and OECD, 2015; for a study involving Italian data see Contini et al., 2017).[3] Beyond averages, attention has been put on gender differences in ability variance, as a large disparity is found at the upper and lower tails of the math performance distribution in both the US (Ellison and Swanson, 2010; Pope and Sydnor, 2010) and Italy (Matteucci and Mignani, 2021). However, the gender gap in math test scores may exaggerate the gender differences in math skills. Economic experiments show indeed that males and females respond differently to competitive test-taking environments

[3] Although it was initially claimed that there are biological differences between males' and females' capacities, this idea was progressively discarded (see the review by Avolio et al., 2020).

(Niederle and Vesterlund, 2010) and females' underperformance increases with competitive pressure (Iriberri and Rey-Biel, 2019). Moreover, several meta-analyses provide evidence that there are no or minimal gender differences in math test results among US primary, middle, and high school students (Hyde et al., 2008; Else-Quest et al., 2010; Lindberg et al., 2010) as well as among OECD students between 15 and 16 years old (Stoet and Geary, 2018), and many argue that the underrepresentation of women in STEM studies is not driven by gender differences in math ability alone (Justman and Méndez, 2018; O'Dea et al., 2018).

Research in educational psychology suggests that self-perception of—or confidence in—math ability plays an important role in explaining gendered educational choices. Nix et al. (2015) and Perez-Felkner et al. (2017), for example, find that US high school boys exhibit higher levels of self-perceived math ability than girls do and that self-perceived math ability is a significant predictor of the likelihood of enrolling and majoring in scientific fields. Some economic studies corroborate these findings (see Anaya et al., 2022, that use administrative US data), whereas others document that gender differences in beliefs about math ability constitute an insignificant part of the gender gap in major choices (see Zafar, 2013, that collects survey data from Northwestern University sophomores).

Along with objective and self-perceived math ability, math-gender stereotypes or biases, i.e., the preconception that females have less math ability than males, have received increasing attention, especially in the psychological literature. Traditional gender stereotypes, held by females and/or embedded in the family and school environment, are found to impair females' math performance,[4] to negatively affect self-perceived math ability, as well as math confidence and attitude,[5] and in turn to shape interest and preferences for the educational track to pursue.[6] Economists have started to analyse the issue in recent years and much economic

[4] In a meta-analysis based on studies with children/adolescents below 18 years of age as participants, Flore and Wicherts (2015) report that girls underperform on math and science tests as a result of the activation of negative math-gender stereotypes and the overall average effect is rather stable over different ages.

[5] In a study involving Italian primary and middle school students, Muzzatti and Agnoli (2007) show that 8th grade girls evaluate themselves less confidently in math when a math-gender stereotype is made salient. In a subsequent study, Passolunghi et al. (2014) find that explicit (but not implicit) math-gender stereotypes are related to self-perception of math ability in Italian 3rd, 5th, and 8th graders.

[6] Davies et al. (2002) find that Canadian female undergraduates exposed to gender-stereotypic television commercials report less interest in quantitative majors and careers than in verbal majors and careers. In a sample of French 6th and 8th graders, a study by Plante et al. (2019) indicates that between-domain stereotypes that males are better than females in math relative to language arts account for the gender gap in math interest.

research has been devoted to the study of teachers' gender stereotypes and their short- and long-run consequences on performance and educational choice (Lavy, 2008; 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 decisions and beliefs in male-typed domains. In an experimental labour market, Reuben et al. (2014) find that without any information other than a candidate's physical appearance, males are more likely to be hired for an arithmetic task than are females, although performance in the task does not differ between genders. Bordalo et al. (2019) show that gender stereotypes distort beliefs about ability of oneself and others, particularly men's beliefs.

However, none of the existing experimental studies has linked experimentally measured individual math-gender stereotypes to upper secondary school choices. This is one of the issues explored in this paper, along with the investigation of further possible determinants of such track choices. In our lab-in-the-field experiment, 8th graders from 5 public middle schools located in Trento (in northeastern Italy) perform an incentivised math task and we use performance in this task (namely the number of correctly solved problems) as a proxy for their math ability. Based on this performance, we elicit incentivised beliefs about own absolute math ability and own relative-to-the-median math ability, which we use to construct measures of confidence. Performance in the math task serves also as the basis for our incentivised measures of individual gender stereotypes that are collected using a matching decision task and a belief task. In the former task, participants are asked to choose whether they would like to be matched with a male or a female 8th grader to compute their earnings in either a competitive or a noncompetitive payoff scheme, depending on the treatment. In the noncompetitive payoff scheme—or *team-up* treatment—participants are paid based on the chosen student's math performance, similar to Reuben et al.'s (2014) experiment. In the competitive scheme—named *competition* treatment—participants are paid based on their own performance only if they did not do worse than the student they chose; otherwise, they earn nothing. The other incentivised measure of individual math-gender stereotype (namely the belief task) requires participants to guess the average math performance of a male and female student. This allows us to test not only whether beliefs are biased by gender, but also whether they are consistent with the gender chosen in the matching task.

Besides assessing students' math-gender stereotypes in an innovative way, via a purposely designed experiment, we also administer more traditional measures of stereotypes, namely the Gender-Science Implicit Association Test (IAT) and survey questions. The IAT is based on timed classification tasks and investigates the presence of implicit, automatic associations between gender and math/science. The survey questions ask for consciously-held beliefs about gender differences in math and thus capture explicit and deliberate stereotyping. Economists have recently studied gender discrimination 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).

The crucial feature of our study is that we analyse how all the experimentally derived variables predict the participants' decision to enrol in a math-intensive versus a non math-intensive high school, information that we collect in the survey together with other pieces of information, such as the teachers' formal track recommendation, the students' perceived value of math, and the interest in math conveyed by the school.

We document a significant gender gap in school choices, with males selecting a math-intensive high school track almost twice as much as females. Although boys and girls do not differ significantly in their overall performance in the math task, highly math-talented boys are more likely to choose a math-intensive high school than highly math-talented girls. We do not identify a consistent and systematic math-gender stereotype in the population under study. Actually, in line with some previous studies, we find that all participants, regardless of gender, believe that females on average outperform males in the math task. As to the factors affecting the decisions to enrol in a math-intensive high school, we observe that they differ between genders. While self-perceived math ability as well as implicit and explicit math-gender stereotypes drive boys' choice of a math-intensive track, for girls a crucial role is played by perceived comparisons with peers. Our results show indeed that girls who believe to be worse, rather than better, than the median in terms of math ability and girls who exhibit math-gender stereotypes in favour of males, as measured by beliefs about others' performance, are less likely to choose a math-intensive high school.

Our paper adds to the economics literature in two main ways. First, this is one of the few papers examining the determinants of female enrolment in upper secondary education. Most previous studies have focused on enrolment and persistence in STEM tertiary education (college and universities), while gender segregation may start developing earlier. In many European and Asian countries, students are tracked into math- 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 choices.

Second, we add to the emerging and promising research in economics 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 major-specific earnings, which are in turn related to college major choices.

The remainder of the article is organised 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,[7] and until then the school curriculum is identical for all students. At the end of the first semester of grade 8, students have to formally enrol into the track of upper secondary education they want to pursue. They make this first track choice by completing an application form together with their parents, aided by teachers' non-binding recommendations.

Upper secondary education 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. In the high school category, there are three different tracks: lyceum, technical high school, and vocational high school. The lyceum offers an academic training and students can choose one of six curricula: humanities, scientific, languages, human sciences, artistic, and music and dance. The technical high school provides an economic curriculum (dedicated to economics and management) and a technological curriculum (specialised in subjects such as technology, informatics, and electronics). 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). Entry into high school is not competitive and does not depend on grades or ability. Thus, the formal application completed by the students and their family guarantees access to the chosen high school. Changing the choice stated in the application is possible, but is costly time-wise and occurs rarely.

In this study, we divide high schools into two categories: STEM and non-STEM. In the STEM high school 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. As a matter of fact, the MIUR last released report about the transition from high school to university indicates that 46.9% (75.1%) of students attending the scientific lyceum (technological curriculum in the technical high school) enrol in STEM universities, while only 12.1% of students attending human sciences lyceum or linguistic lyceum do so.

[7] At the beginning of grade 6 pupils are allocated to classes of 18–27 students and stay in the same class for all three years. Classes are formed randomly within schools, respecting the criterion of equal allocation of students across classes according to gender, socioeconomic status, and ability level (as reported by the primary school teachers).

3 The study design

3.1 Data collection

A total of 548 Italian 8th graders from 27 different classes participated in the study, which took place in May 2018. They were recruited from 5 public middle schools in Trento in Northeast Italy.[8] 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.

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 and data collection lasted about 60-70 minutes per class. In each class, prior to data collection, each student received a unique ID number that was sellotaped to her desk and printed on any document she received. Written general instructions were distributed and read aloud by an experimenter. The general instructions emphasised the importance of no communication among participants and the anonymity of participation. They also illustrated the structure of the study and the procedure for rewarding.[9]

The variables of interest were collected in five different parts. We first assessed each student's objective math ability (part 1).[10] Then, we collected incentivised measures of individual math-gender stereotypes and math confidence (parts 2 and 3). Next, we investigated the existence of implicit gender stereotypes (part 4). Finally, we administered the survey (part 5). Although student participants knew from the beginning that there would be five parts, they learned about the content of each part only after having completed the previous one.

As we will detail in the following subsection 3.2, students could earn points for their choices in parts 1 to 3. 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 at the Trento Mondadori bookstore. Only three

[8] Students were admitted to the study after having obtained formal permission from the dean of their school and informed consent from their parents.

[9] A translation of the full set of instructions (originally in Italian) can be found in Appendix A.

[10] Math ability was measured at the beginning of the study to avoid any potential influence of gender priming on performance.

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 part 5 by placing all ID numbers in a box and by asking a volunteer to pick three numbers from the box. Participants did not receive any information about their own experimental earnings or the experimental earnings of others during the experiment. Only when the three randomly selected participants received their vouchers, everyone was informed about the earned points. Average earnings (computed over all participants) amounted to €8.6, with a minimum of zero and a maximum of €21.

The five schools in our sample are representative of the general population of Italian 8th graders in public schools.[11] First, as we will discuss later, the high school track choices by gender observed in our sample are 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; 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%, where the national average is 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.

Math ability. The first task that student participants performed was a math task, which consisted of ten multiple-choice and open-ended problems taken, in a modified version, from the INVALSI (National Institute for the Evaluation of the Italian Education System) math tests for grade 8.[12] Students had 15 minutes to answer as many problems as they could.

[11] In Italy there are 8064 middle schools, 7418 of which are public.

[12] These are Italian standardised assessment tests administered every year to students attending grades 2, 5, 8, and 10 to evaluate their proficiency in mathematics, Italian, and English. We chose the INVALSI tests because the math performance of girls and boys is not significantly different in the considered macro-

They were not allowed to use calculators, but could use the blank space next to each problem to make computations. There were always 7 distinct versions of the ten problems to avoid copying from neighbouring students. 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.

Confidence: absolute and relative beliefs about own math ability. The high school track choice 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. We gave participants a financial incentive to report beliefs accurately: we paid them 1 point for each belief that turned out to be correct.

Incentivised measures of individual math-gender stereotypes. The main aim of the experiment is to obtain incentivised measures of individual math-gender stereotypes, which are collected in a matching decision task and in a belief elicitation task.

In the matching decision task, participants were asked to refer to the math task performed in part 1 and to decide which one of two randomly selected 8th graders—called A and B—they wanted to be matched with for determination of their earnings. To avoid making gender overly salient, we asked participants to make two such decisions. In Decision 1, the choice was between a male and a female student, who were classmates in a different participating school; which gender was presented first was randomised. In Decision 2, the choice was between a student in a different class of the same school and a student in a different school. We will analyse data only from Decision 1 as Decision 2 serves the purpose to divert participants' attention from gender.

The rule determining the earnings depends on the treatment a participant is assigned to. In the *team-up* (TU) treatment, echoing a hiring task, participants earned 1 point for

geographical area (North-East Italy), even though gender differences are observed at the national level.

each correct answer given in the math task by the chosen partner. In the *competition* (CM) treatment, participants earned 1 point for each correct answer they gave in the math task if, in this task, they had performed overall not worse than the chosen partner; otherwise, they earned 0 point. The two treatments were implemented in a between-subjects design. We ran the CM treatment as a robustness check to test whether making decisions in a framework involving a tournament affects the choice of the partner (male or female) with whom one wants to be matched with.

In both treatments, participants made each decision by dividing a total of 10 virtual balls between A and B. We marked the balls as indicated by each participant, put them in a box, and drew one at random. The letter on the drawn ball determined the matched partner. Participants were reminded that the higher the number of balls allocated to one of the two students, the higher the probability that this student would be matched with them.[13]

In the belief elicitation task, participants were asked to guess the number of correct answers given, on average, by A and B in the math task. Each correct guess was rewarded with 1 point. The beliefs concerning the male and female student's performance (namely the beliefs referring to Decision 1) serve as a further experimental measure of math-gender stereotypes.

Implicit math-gender stereotypes. In addition to the incentivised measures of individual stereotypes, we collected unincentivised measures of implicit gender stereotypes using the gender-math/science Implicit Association Test (IAT) (Greenwald et al., 1998). In the paper-and-pencil format test 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

[13] We let participants freely choose any distribution of balls to allow for coarseness of preferences, although to maximise expected outcome a person, when not indifferent between alternatives, should distribute the ten balls as (10, 0) or (0, 10).

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 randomised across participants. For each sheet, participants were given 45 seconds to categorise as many words as possible starting from the top of the list without skipping any word or correcting mistakes.[14] A measure of implicit math-gender stereotypes is obtained by comparing how correctly participants categorise the words in the two sheets.

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 5-point Likert scale from "girls are definitely better" to "boys are definitely better".

High school track choice and other survey measures related to math. In the survey, student participants were asked to report the chosen high school track as well as the official teachers' recommendation. We also collected three math-related measures in the survey: (i) the participants' degree of confidence in their math skills, (ii) their perceived value of math, and (iii) the interest in math conveyed by the school. Students' math 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 is composed of 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.[15] 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 1) be aware of the applications of math in real life, 2) realise how math is relevant for

[14] Participants were allowed to familiarise 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.

[15] See part 5 of the instructions in Appendix A for the statements composing the TIMSS scale and the Lim and Chapman's subscale.

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

everyday decisions, 3) get excited about math.

4 Results

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

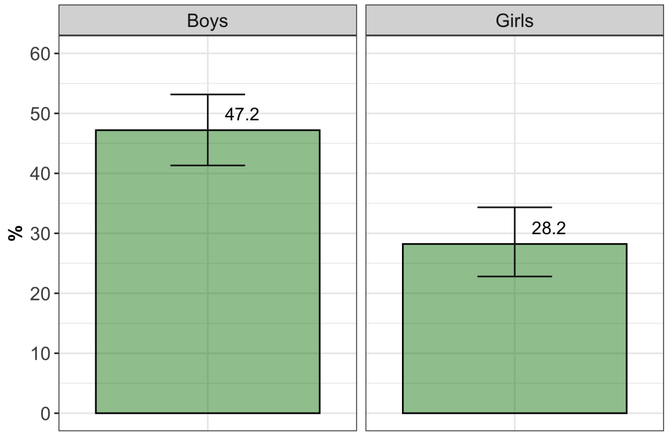
4.1 The high school track choice

The percentages of female and male students who opted for a STEM high school in our sample are reported in Figure 1. Almost half of the boys chose a STEM high school, while only 28.2% of girls did so. A test of proportions indicates that boys are significantly more likely than girls to choose a STEM high school (Pearson χ^2 test p-value<0.001).[16] The gender difference in high school track choice observed in our sample mirrors that observed 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.[17]

In the following analysis, we investigate what variables can help explain this lopsided choice. For the sake of brevity, we will sometimes refer to the females (males) who opted for a STEM high school as "STEM girls" ("STEM boys"). Similarly, "non-STEM girls" ("non-STEM boys") will stand for females (males) who chose a non-STEM high school.

[16] All statistical tests presented in the paper are two-sided.
[17] See `http://dati.istruzione.it/espscu/index.html?area=anagStu`.

Figure 1: Frequency of STEM high school choices by gender



Notes. The height of the bars and the number above the bars indicate the percentage of boys and girls that chose a STEM high school. The figure also reports confidence intervals from Pearson χ^2 test.

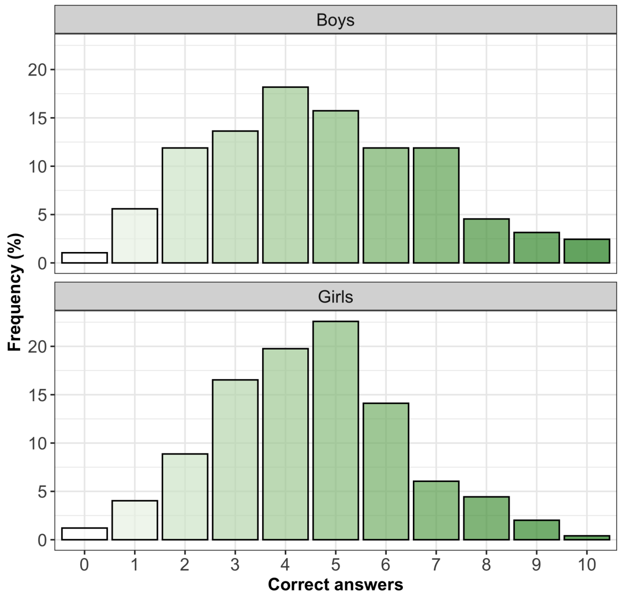
4.2 Math ability and high school track choice

In this section we assess whether there are gender differences in our measures of math ability and analyse whether objective and self-perceived math ability can account for the observed gender difference in the high school track choice.

4.2.1 Objective math ability

Figure 2 presents the distribution of correct answers in the math task for boys and girls separately as well as the corresponding descriptive statistics. 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 for both genders. The answers of boys are slightly more dispersed than those of girls, as evident from the standard deviation. In terms of symmetry, both distributions are positively skewed with more mass 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, henceforth WRS, test 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. Dividing the sample into subsamples based on the chosen type of high school, we find that the students who

Figure 2: Math performance by gender



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

Notes. The figure portrays the distribution of correct answers in the math task separately for male and female students. The table reports descriptive statistics of the distributions (SD stands for standard deviation).

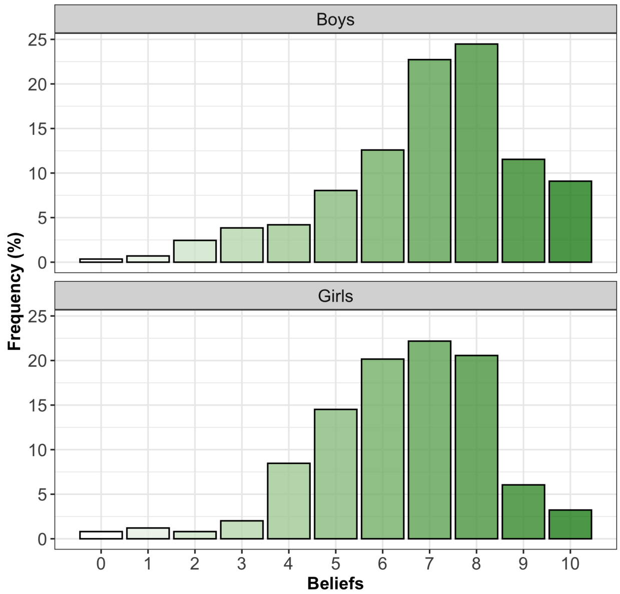
chose a STEM high school significantly outperform those who chose a non-STEM high school and this holds for both girls (WRS test p-value=0.005) and boys (WRS test p-value<0.001).

4.2.2 Beliefs about own math ability

Figure 3 depicts the distribution of the participants' beliefs about their own math performance, separated by gender, displaying descriptive statistics in tabular form. The median equals 7 for both genders, but the mean estimate is higher for male students. The figure makes it apparent that boys' beliefs are more concentrated on the high values of the 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 p-value< 0.001).[18] Considering the high school track choice, STEM boys tend to be

[18] 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)

Figure 3: Expected performance in the math task by gender



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

Notes. The figure portrays the distribution of expected correct answers in the math task separately for male and female students. The table reports descriptive statistics of the distributions (SD stands for standard deviation).

more confident in their math ability than non-STEM boys (WRS test p-value< 0.001). In contrast, STEM and non-STEM girls do not differ significantly in this kind of confidence (WRS test p-value=0.106).

Turning to the second measure of confidence that we elicited, i.e., the perceived position relative to the median, Table 2 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 (Pearson χ^2 test p-value<0.001), with most boys believing to be better than the median and

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 males than for females (WRS test p-value=0.071). Additionally, while non-STEM boys overestimate their performance more than STEM boys (WRS test p-value= 0.002), no statistically significant difference is found between STEM and non-STEM girls (WRS test p-value= 0.298).

Table 2: 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%

Notes. The table reports the percentages of male and female students who perceive their performance in the math task as worse than, the same as, and better than the median performance of their classmates.

most girls believing to be as good as the median.[19] When comparing the distributions for a given gender across track choices, we find that 42.2% (21.0%) of STEM (non-STEM) girls think they are better than the median (Pearson χ^2 test p-value=0.001) and 65.8% (36.0%) of STEM (non-STEM) boys think so (Pearson χ^2 test p-value< 0.001).

4.2.3 Relationship between incentivised measures of math ability and high school choice

To precisely assess whether objective and self-perceived math ability relates to the high school track choice, we estimate a Generalized Linear Model (logit regression) where the dependent variable is equal to 1 (0) if the student chose (did not choose) a math-intensive high school track. Table 3 reports the estimation results, using fixed effects at the class level. We present distinct estimates for girls and boys to ease the interpretation of the coefficients.[20]

The independent variables in columns (1) and (3) refer to, respectively, girls' and boys' objective math ability categorised in quartiles of performance in the math test ('MATH Q1' is associated to the lowest quartile and 'MATH Q4' to the highest quartile). The regression outcomes show that objective math ability has different predictive power for girls and boys. Boys in higher performance quartiles are more likely to enrol in a STEM high school. Even

[19] 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 overplacement, 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 overplace their relative performance (37.8% vs 33.8%) and more girls than boys underplace it (25.1% vs 15.6%). The gender difference in underplacement is statistically significant (Pearson χ^2 test p-value=0.033). Dividing the sample according to the high school track choice, we find that the beliefs on the relative performance are (i) more accurate for STEM than non-STEM boys (Pearson χ^2 test p-value< 0.001) and (ii) not significantly different between STEM and non-STEM girls (Pearson χ^2 test p-value$= 0.554$).

[20] The difference in the number of observations across specifications is due to missing values in the explanatory variables (see Appendix B for a few summary statistics of the variables).

Table 3: Determinants of the high school track choice

	Girls		Boys	
	(1)	(2)	(3)	(4)
(Intercept)	-2.142 (0.775)**	-3.590 (1.275)**	-0.638 (0.814)	-2.181 (1.111)*
MATH Q2	0.927 (0.517)$^{\circ}$	1.527 (0.596)*	0.923 (0.454)*	0.589 (0.487)
MATH Q3	1.105 (0.476)*	1.354 (0.555)*	1.847 (0.423)***	1.565 (0.459)***
MATH Q4	1.151 (0.594)$^{\circ}$	1.019 (0.711)	3.236 (0.522)***	2.689 (0.591)***
MATH belief		0.145 (0.139)		0.218 (0.104)*
MATH median		1.233 (0.490)*		0.265 (0.364)
AIC	285.174	247.253	343.981	311.846
BIC	390.576	350.526	453.661	422.465
Log Likelihood	-112.587	-93.627	-141.991	-124.923
Num. obs.	248	231	286	262

Notes. Coefficients are from logit regressions (GLM), where the dependent variable equals 1 if the student chose a STEM track. All regressions have class fixed effects. Symbols ***, **, *, and $^{\circ}$ denote statistical significance at the 0.1%, 1%, 5%, and 10%, respectively.

though for girls there is a significant difference between the lowest and the other quartiles, the estimated parameters are very similar for distinct quartiles. A series of linear hypothesis tests confirms that the estimated parameters for distinct quartiles are significantly different for boys (χ^2, all p-values$<$0.05), but not for girls (χ^2, all p-values$\geq$0.696). Importantly, including other control variables makes the coefficient of 'MATH Q4' insignificant in the girls' regressions, implying that best female performers are not more likely to choose a STEM high school than worst female performers. Thus, we can state:

Result 1 *Objectively measured math ability is more likely to affect the STEM high school choice of boys than that of girls.*

Columns (2) and (4) add students' beliefs about their absolute and relative performance to the previous specification. The variable 'MATH belief' refers to a student's predicted number of correct answers in the math test. 'MATH median' is a dichotomous variable taking value 1 if a student believes to be strictly above the class median in the math test. Controlling for actual performance, the two types of elicited beliefs influence the likelihood of choosing a math-intensive high school track differently for girls and boys. The coefficient of 'MATH median' is positive and statistically significant for girls (but not for boys), meaning that girls who believe to be above the class median are more likely to choose a STEM high school.

For boys, it is the coefficient of 'MATH belief' that is positive and statistically significant, meaning that the higher the expected number of correct answers, the more likely it is that boys choose a STEM high school. We summarise the results on the relationship between the incentivised measures of self-perceived math ability and STEM high school choices as follows:

Result 2 *Girls are significantly more likely to enrol in a STEM high school 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.*

4.2.4 Survey measures related to math

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 then over students, we obtain a mean confidence score of 3.044 for girls and 3.330 for boys.[21] Both girls and boys display an overall moderate confidence in math. Yet, compared to boys, girls feel significantly less confident about their math ability (WRS test p-value=0.003). When considering the chosen high school track, we observe that STEM girls and STEM boys are significantly more confident in math than their non-STEM counterparts (WRS test p-value<0.001 for both comparisons).

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 being valuable, with mean scores respectively of 3.816 and 3.902, and there is no gender difference in the value attributed to math (WRS test p-value=0.183). If we further divide the sample by type of chosen high school, compared to non-STEM girls (boys), STEM girls (boys) perceive math as significantly more valuable (WRS test p-value<0.001 for both comparisons).

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

[21] Statements expressing negative sentiments are reverse coded.

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). Considering the chosen type of high school, both STEM girls and STEM boys perceive math courses as a greater boost to their math interest than their non-STEM counterparts (WRS test p-value<0.001 for both comparisons).

To summarise, the math-related (non-incentivised) survey measures confirm our previous results based on incentivised measures of confidence—namely that girls are, on average, less confident than boys in their math ability—and additionally indicate that both genders perceive (i) math as valuable and (ii) the school as important to the development of their interest in math, although girls report a significantly lower score on the questions related to the school.

4.3 Individual math-gender stereotypes and high school track choice

Proceeding as in Section 4.2, we first test whether female and male participants hold individual math-gender stereotypes as measured by the incentivised tasks, the IAT, and the survey questions. Then, we examine whether our measures of individual math-gender stereotypes help explain the observed gender difference in the high school track choice.

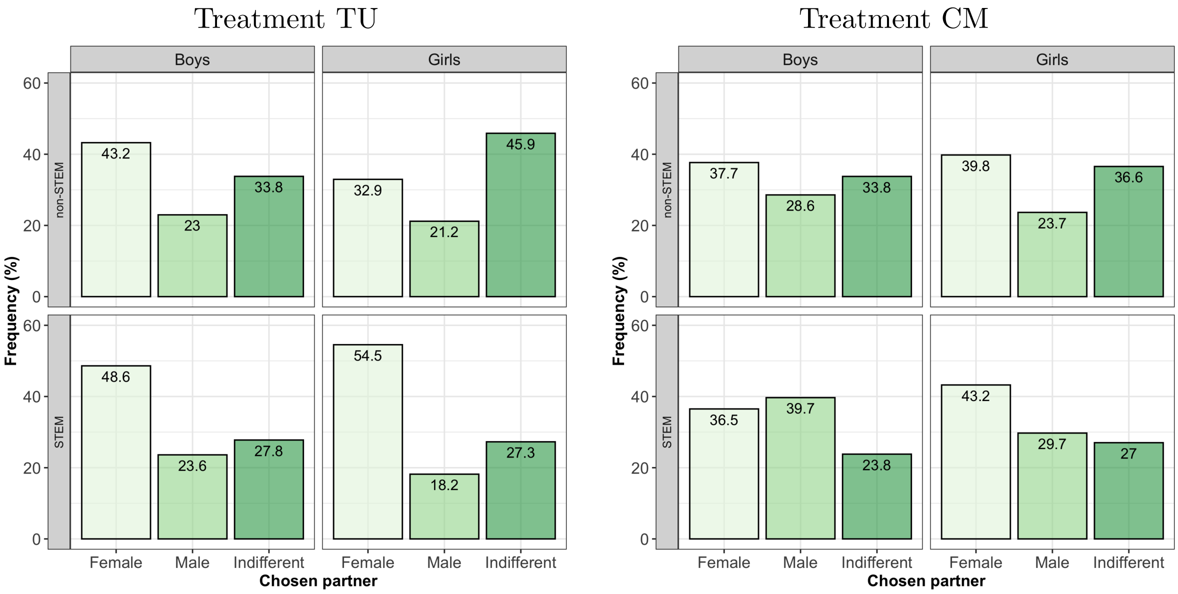
4.3.1 Incentivised measures of individual math-gender stereotypes

We obtained incentivised measures of individual math-gender stereotypes in a matching decision task under two treatments and in a belief elicitation task.

We start by analysing decisions in the matching decision task.[22] In treatment TU, most participants are found either to prefer being matched with a female student or to be indifferent to the partner's gender. Specifically, 39.0% (45.9%) of girls (boys) assign more balls to a female student and 40.7% (30.8%) of girls (boys) divide the balls evenly between the male

[22] In both the TU and CM treatments, preferences for the partner's gender were expressed by distributing a total of 10 balls between the male and the female student. In the analysis, we assume that a partner is preferred if the number of balls assigned to him or her is greater than 5. Although expected outcome maximisation predicts that participants, when not indifferent, should assign all balls to the preferred option, the empirical distributions indicate that this is not always the case, possibly reflecting coarse preferences.

Figure 4: Preferences for female and male partners in the matching decision task by treatment, gender, and high school track choice



Notes. In each graph, the height of the bar and the number within the bar indicate the percentages of participants in the corresponding subgroup that assign more balls to a female ('Female' bar), more balls to a male ('Male' bar), and split the 10 balls equally between the female and the male ('Indifferent' bar).

and the female student. There are no significant differences in preferences for the partner's gender between girls and boys (Pearson χ^2 test p-value=0.248). The left panel of Figure 4 depicts the frequencies with which participants in the TU treatment prefer, or are indifferent to, being matched with a female or a male student separately for each gender and each chosen high school track. In all subgroups, males are the least preferred partners and every subgroup, except non-STEM girls, exhibits a bias in favour of females.[23] This indicates that our participants do not hold, on average, traditional gender-stereotyped preferences when earnings depend on the partner's math performance.

In treatment CM, where participants earn points only if their performance in the math task is not worse than that of the chosen partner, math-gender stereotypes become apparent because participants, regardless of their gender, prefer competing against a female student. We find indeed that 40.8% (37.1%) of girls (boys) assign more balls to a female student and

[23] The results of Pearson χ^2 tests show a marginally significant difference between STEM girls and non-STEM girls (p-value=0.083), but not between STEM boys and non-STEM boys (p-value=0.718).

25.4% (33.6%) of girls (boys) assign more balls to a male student, and that the proportions of balls allocated to a female by the two genders are not significantly different (Pearson χ^2 test p-value=0.333). The preference for being matched with a female student is generally detected also if we further split the sample according to the high school track choice. This can be seen in the right panel of Figure 4, which shows that in all but one subgroup (i.e., STEM boys), participants assign more balls to a female student.[24]

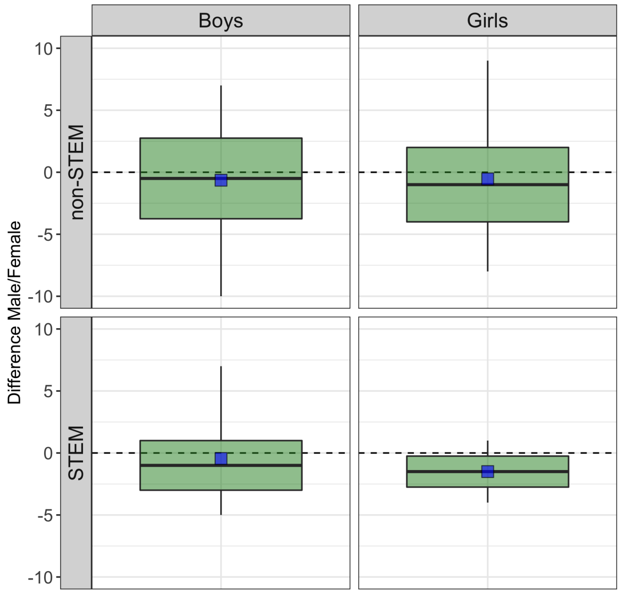
Turning to the measure of math-gender stereotypes collected in the belief elicitation task, our results show that both girls and boys expect, on average, a female student to have performed better in the math task than a male student, with the expected mean difference in math performance between the male and the female student being different from zero (WSR test p-values<0.001 for both genders) and equal to -0.423 for girls and -0.472 and boys. Furthermore, the beliefs of girls and boys are not significantly different (WRS test p-value=0.687).[25] Figure 5 displays box plots of the expected difference between the male student's performance and the female student's performance separately for each gender and each chosen high school track. The median and mean values are negative for all four subgroups, which indicates a consistent bias in favour of females (WSR test p-values<0.001 for all comparisons). Comparing the bias between subgroups by means of WRS tests reveals that the STEM girls' bias in favour of females is significantly higher than that of non-STEM girls (p-value=0.026), while no significant difference is registered between STEM and non-STEM boys (p-value=0.271).

Recalling that all participants, regardless of their gender, prefer being paid based on a female student's performance in treatment TU and competing against a female student in treatment CM, the elicited beliefs are in line with choices in TU, but are not so in CM. To more precisely understand how the choice of the partner in the two experimental treatments relates to his/her expected math performance, we estimate OLS regressions, separately for girls and boys, where the dependent variable is the difference between balls allocated to a male

[24] The preferences of STEM girls do not differ significantly from those of non-STEM girls and the same holds for the comparison between STEM boys and non-STEM boys (the respective Pearson χ^2 test p-values are 0.556 and 0.292).

[25] Because beliefs about the others' math performance should be independent of the treatment, we pool these belief data from the two treatments. It is however worth noting that participants in both TU and CM hold on average beliefs favouring females (WSR test p-values<0.001 for both treatments).

Figure 5: Expected difference between a male's math performance and a female's math performance by gender and high school track choice



Notes. The box plot in each panel refers to a different subgroup and provides a visual summary of the distribution of the expected difference in math performance between the male and the female student. The filled squares indicate mean values.

and balls allocated to a female in the matching decision task, and the independent variables are the expected difference in math performance between the male and the female student ('Diff Exp Performance'), a treatment dummy ('TU'), and a term for the interaction between these two variables. Table 4 presents the estimates of our regressions. The coefficient of the interaction term is positive and statistically significant for girls, indicating that girls who believe that a male student has performed better (worse) than a female student in the math task tend to allocate more (less) balls to a male in treatment TU. No significant impact of beliefs on allocated balls is detected for girls in treatment CM and for boys in both treatments. Additionally, the results of the boys' regression show that boys are significantly less likely to pick a male in TU than in CM, irrespective of beliefs. This is consistent with our previous finding that most boys (86.6%) in TU either assign more balls to a female student or divide the balls evenly between students, whereas this occurs less often (66.4% of the times) in CM.

Table 4: Decisions in the matching task and beliefs about males' and females' math performance

	Girls	Boys
Intercept	−0.270 (0.395)	0.489 (0.453)
Diff Exp Performance	0.125 (0.248)	0.085 (0.246)
TU	0.280 (0.589)	−1.339 (0.613)*
Diff Exp Performance × TU	0.983 (0.376)**	0.136 (0.338)
R^2	0.061	0.022
Adj. R^2	0.050	0.011
No. obs.	248	286

Notes. Coefficients are from OLS regressions, where the outcome variable is the difference between balls allocated to males and balls allocated to females in the matching decision task. Robust standard errors are reported in parentheses. ** and * denote statistical significance at the 1% and 5%, respectively.

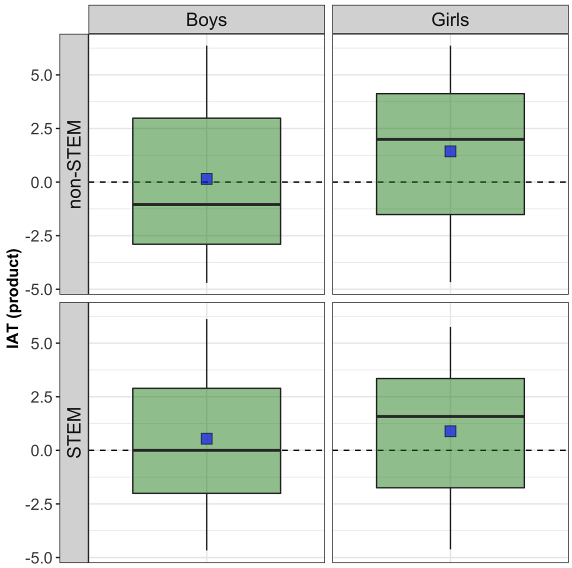
4.3.2 Individual math-gender stereotypes at the implicit level

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).[26] Larger scores denote larger implicit math-gender stereotypes, i.e., more 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.

Figure 6 draws, separately for each gender and each chosen high school track, box plots of the distribution of IAT scores. Compared to non-STEM girls, STEM girls do not show lower implicit stereotypes (WRS test p-value=0.301). Conversely, STEM boys display significantly higher levels of implicit stereotypes than non-STEM boys (WRS test p-value=0.004).

[26] 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 categorisations in the stereotype-congruent condition and B as the number of correct categorisations in the stereotype-incongruent condition. Then, we define $X = max\{A, B\}$ and $Y = min\{A, B\}$. The IAT score is calculated as $\sigma = \frac{X}{Y} \times \sqrt{X - Y}$. 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 categorisations. However, results do not change when the entire sample is considered.

Figure 6: Implicit math-gender stereotypes by gender and high school track choice



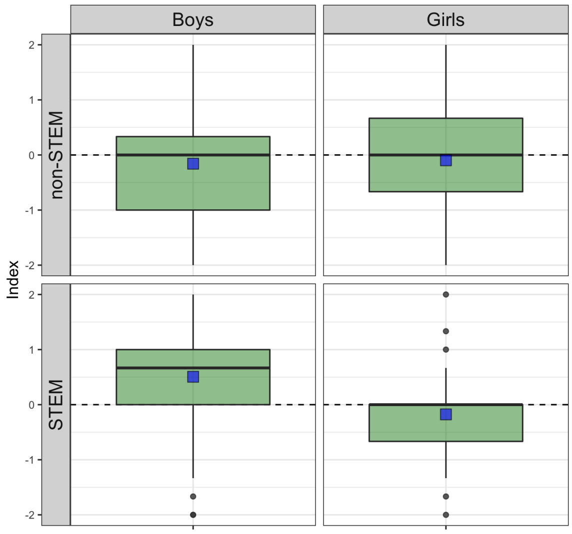
Notes. The box plot in each panel refers to a separate subgroup and provides a visual summary of the distribution of that subgroup's IAT scores. The filled squares indicate mean values.

4.3.3 Individual math-gender stereotypes at the explicit level

To assess the participants' explicit endorsement of math-gender stereotypes, we collected responses to three survey questions regarding the students' own beliefs about gender differences in math and their perception of their teachers' and classmates' math-gender stereotypes. In the analysis, all answers—from 0 to 5—are recoded from -2 to 2, so that the scale midpoint is at 0 and positive scores imply agreement with the belief that males are better at math. Since the internal consistency of the three answers is satisfactory (Cronbach's alpha = 0.86), we will take the average of the three answers, which we call *explicit bias index*, as primary unit of analysis. The mean index equals -0.119 for girls and 0.157 for boys; both values are significantly different from zero (the respective WSR test p-values are 0.034 and 0.016). Thus, both genders show a moderate explicit bias favouring their gender.

Figure 7 displays the distribution of individual-level scores obtained for the explicit bias index, separated by gender and chosen high school track. The mean value of the index is slightly negative—and the median equals 0 (indicating no explicit bias)—for all subgroups except STEM boys. When testing for differences between subgroups, we find a statistically

Figure 7: Explicit bias index by gender and high school track choice



Notes. The box plot in each panel refers to a separate subgroup and provides a visual summary of the distribution of that subgroup's explicit bias index. The filled squares indicate mean values.

significant difference for STEM vs non-STEM boys (WRS test p-value< 0.001), but not for STEM vs non-STEM girls (WRS test p-value=0.464).

4.3.4 Relationship between individual math-gender stereotypes and high school choice

We turn to evaluate whether high school track choices are related to our measures of individual math-gender stereotypes, controlling for objective and subjective math ability variables. To this aim, in Table 5 we add four independent variables to Table 3 regression. The variable 'Matching decision' captures choices in the two experimental treatments of the matching decision task. To let positive values of the variable represent traditional math-gender stereotypes, we define it as the difference between balls assigned to males (females) and balls assigned to females (males) for TU (CM). 'Diff Exp Performance' denotes, as before, the participants' expected difference in math performance between male and female students. 'Own implicit bias' is the participants' IAT scores. Finally, 'Explicit bias index' is the index of explicit math-gender stereotypes defined in Section 4.3.3.

Table 5: Determinants of the high school track choice (continued from Table 3)

	Girls	Boys
(Intercept)	-3.952 (1.621)*	-4.276 (1.429)**
MATH Q2	1.626 (0.653)*	0.127 (0.546)
MATH Q3	1.459 (0.616)*	1.319 (0.538)*
MATH Q4	1.119 (0.777)	3.079 (0.716)***
MATH median	1.432 (0.521)**	-0.047 (0.423)
MATH belief	0.179 (0.151)	0.361 (0.131)**
Matching decision	-0.018 (0.051)	0.040 (0.039)
Diff Exp Performance	-0.551 (0.206)**	-0.081 (0.110)
Own implicit bias	-0.058 (0.047)	0.126 (0.047)**
Explicit bias index	0.075 (0.295)	0.587 (0.222)**
AIC	239.530	274.281
BIC	354.603	395.958
Log Likelihood	-85.765	-102.141
Num. obs.	218	239

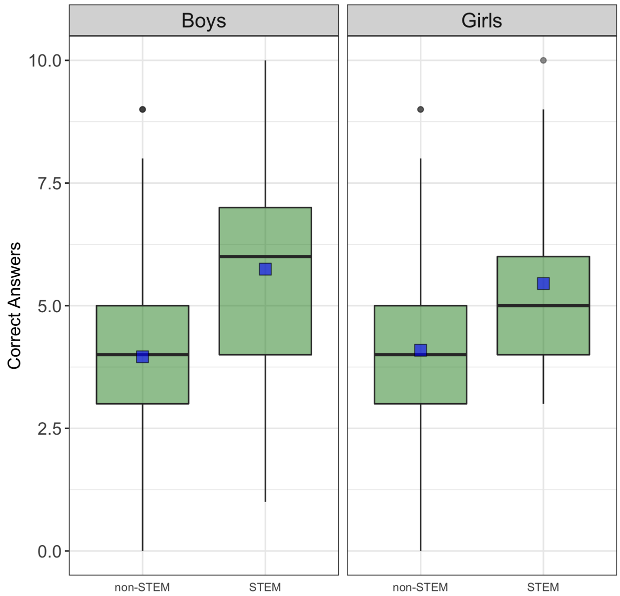
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes. Coefficients are from logit regressions (GLM), where the dependent variable equals 1 if the student chose a STEM track. All regressions have class fixed effects. Symbols ***, **, *, and $^{\circ}$ denote statistical significance at the 0.1%, 1%, 5%, and 10%, respectively.

Our main findings are as follows. First, looking at the estimates for girls, the coefficient of 'Diff Exp Performance' is negative and statistically significant, meaning that girls who believe that males outperformed females in the math task are less likely to enrol in a STEM high school. Our further incentivised measure of math-gender stereotypes does not significantly impact the girls' likelihood of choosing a STEM high school. Second, examining the boys' regression, the coefficients of 'Own implicit bias' and 'Explicit bias index' are positive and statistically significant. Neither of our incentivised measures of stereotypes is significantly related with the boys' high school track choice. We can therefore state the following result:

Result 3 *Girls holding more gender stereotyped (incentivised) beliefs about math ability are significantly less likely to enrol in a STEM high school. For boys, stronger implicit and explicit math-gender stereotypes are significantly associated with higher likelihood of choosing a STEM high school.*

Figure 8: Math performance by gender and track (STEM vs non-STEM) recommended by teachers



Notes. The box plots show the distributions of correct answers in the math task for each gender, grouped by teachers' track recommendation.

4.4 Teachers' recommendations

We conclude our analysis by examining the potential math-gender stereotypes held by the teachers by considering their official high-school track recommendations, controlling for math ability. Figure 8 depicts, for the track (non-STEM vs STEM) recommended by the teachers, the distribution of correct answers in the math task of boys and girls. Teachers tend to recommend a STEM high school to the students with higher math performance, whatever their gender. WRS tests confirm that the difference in math performance between students recommended for STEM track and students recommended for non-STEM track is statistically significant for both boys and girls (both p-values<0.001). We can thus state that the teachers' official track recommendations discriminate not on the basis of gender, but rather on the basis of objective math ability.

5 Discussion and concluding remarks

The reason why girls do not choose the most rewarding STEM educational tracks in terms of future wages and labour 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 high school track choice of Italian 8th graders. 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 choices.

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 in math test results. Despite this equality, girls in our sample are significantly less likely to choose a math-intensive high school track.

The main result of our regression analysis is that, controlling for math ability, girls' choice of a STEM high school track hinges on how girls perceive themselves in comparison to others. We find indeed that girls are more likely to choose a STEM track the more they think they are better than the median classmate in terms of math ability and the stronger are their counter-stereotypical (incentivised) beliefs that females on average outperform males in the math task. Neither objective math ability (measured by the number of correct answers in the math task) nor self-perceived math ability (namely the expected number of correct answers in the math task) are found to affect the high school choice of girls, whereas they are significant predictors of the choice of boys, along with implicit and explicit math-gender stereotypes. The observation that beliefs about relative math performance are relevant for the high school track choice of girls (but not of boys) and beliefs about absolute math performance are relevant for the track choice of boys (but not of girls) is consistent with psychological theories suggesting that girls and boys frame their math ability in different ways and focus their attention on distinct aspects when making educational choices (Loyalka et al., 2017).

Our analysis also reveals that the incentivised measures of individual math-gender stereotypes collected in the matching decision task, under either a noncompetitive payoff scheme (the team-up treatment) or a competitive payoff scheme (the competition treatment), do not have a significant effect on high school track choice, neither for girls nor for boys. As we show, the two used payoff schemes actually give rise to different results regarding traditional math-gender stereotypes: 8th graders in our sample prefer being matched with a female student both when their payment depends on the chosen student's math performance (suggesting in this case the absence of stereotypes) and when they earn points only if they did not perform worse than the chosen student (indicating in this case the endorsement of stereotypes). The presence in our sample of counter-stereotypical beliefs is consistent with the stated preferences for the partner's gender in the former case, but not in the latter. Before turning to a discussion of how these apparently inconsistent findings may be explained, it is worth noticing their methodological significance for researchers in general and experimentalists in particular since they point to the importance of performing robustness checks via the implementation of alternative and distinct setups when eliciting a certain variable or testing a specific hypothesis.

While we are unable to conclusively identify the reason for the observed inconsistency between counter-stereotypical beliefs and preferences for a female partner in the competition treatment, such a finding could be accounted for if our participants expect female math performance to exhibit higher variance and more volatility than male performance. Our student participants may indeed think that when girls are bad (good) at math, they are extremely bad (good), although they expect female students to outperform male students on average. In this sense, choosing to be matched with a female rather than a male student may be perceived as a riskier option. It is recognised that competition affects risk-related choices and most previous studies, mainly in the behavioural finance literature (e.g., Gärling et al., 2021), indicate that competition increases risk-taking. Thus, when our student participants are put in a competitive setting where they earn zero if the chosen peer's math performance exceeds their own performance, they may be prone to make risky decisions and choose to be matched with the partner that is perceived as riskier, namely a female student.

A second possible explanation for the conflicting evidence regarding beliefs and expressed preferences for the partner's gender in the competition treatment is based on psychological studies indicating that a competitive relationship in which the individual performance may be rewarded brings about a shift towards increased focus on one's own task, disregarding the other's actual or expected performance (Ruissen and de Bruijn, 2016). Hence, our student participants choose the gender of the partner that they believe has performed better—i.e., a female student—in the team-up treatment, but the shift in focus under competition leads them to neglect their beliefs and focus exclusively on themselves. While this argument may account for the observed inconsistency, it does not provide a justification for the preferences for being matched with a female rather than a male student in the competition treatment.

As to the non-incentivised measures of individual math-gender stereotypes collected in our experiment, we find that girls (but not boys) display automatic gender stereotypical associations, although both genders show 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, is in line with previous evidence on Italian students (Galdi et al., 2014; Passolunghi et al., 2014) and corroborates the hypothesis that implicit math-gender stereotypes develop earlier than explicit ones and that the two types of stereotypes are not necessarily correlated with each other (Passolunghi et al., 2014; Vuletich et al., 2020).

While previous literature has documented that math teachers' gender stereotypes influence the high school track choice and induce girls to attend less math-intensive high schools (Carlana, 2019), in our context teachers recommend STEM tracks to the students with better math performance, without discriminating on the basis of gender.

Overall, our study provides evidence that perceived comparisons with peers are likely to affect girls' educational track choices. 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 to this purpose. Another option (often brought up in the literature) 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 performance. 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 characterised 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.

References

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics*, 100(5): 876–890.

Anaya, L., Stafford, F., and Zamarro, G. (2022). Gender gaps in math performance, perceived mathematical ability and college STEM education: the role of parental occupation. *Education Economics*, 30(2): 113–128.

Avolio, B., Chávez, J., and Vílchez-Román, C. (2020). Factors that contribute to the underrepresentation of women in science careers worldwide: a literature review. *Social Psychology of Education*, 23(3): 773–794.

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

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3): 739–773.

Breda, T. and Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science*, 353(6298): 474–478.

Breda, T. and Ly, S.T. (2015). Professors in core science are not always biased against women: evidence from France. *American Economic Journal: Applied Economics*, 7(4): 53–75.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129(3): 1409–1447.

Buser, T., Peter, N., and Wolter, S.C. (2017). Gender, competitiveness, and study choices in high school: evidence from Switzerland. *American Economic Review: Papers & Proceedings*, 107(5): 125–130.

Caprile, M., Palmn, R., Sanz, P., and Dente, G. (2015). Encouraging STEM studies: labour market situation and comparison of practices targeted at young people in different Member States. European Union, Brussels.

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

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

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

Davies, P.G., Spencer, S.J., Quinn, D.M., and Gerhardstein, R. (2002). Consuming images: how television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12): 1615–1628.

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.

Ellison, G. and Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24(2): 109–128.

Else-Quest, N.M., Hyde, J.S., and Linn, M.C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, 136(1): 103–127.

Flore, P.C. and 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. and Parey, M. (2018). Early gender gaps among university graduates. *European Economic Review*, 109: 63–82.

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

Gärling, T., Fang, D., Holmen, M., and Michaelsen, P. (2021). Financial risk-taking related to individual risk preference, social comparison and competition. *Review of Behavioral Finance*, 13(2): 125–140.

Galdi, S., Cadinu, M., and Tomasetto, C. (2014). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Development*, 85(1): 250–263.

Granato, S. (2020). Early influences and the gender gap in STEM. Available at SSRN: `https://ssrn.com/abstract=3845391`.

Greenwald, A.G., McGhee, D.E., and Schwartz, J. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology*, 74: 14641480.

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

Iriberri, N. and Rey-Biel, P. (2019). Competitive pressure widens the gender gap in performance: evidence from a two-stage competition in mathematics. *The Economic Journal*, 129(620): 1863–1893.

Justman, M. and Méndez, S.J. (2018). Gendered choices of STEM subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of Education Review*, 64: 282–297.

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

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92: 2083–2105.

Lavy, V. and 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., and Nosek, B. (2008). Assessing implicit cognitions with a paper-format implicit association test, in M. Morrison and T. Morrison (eds.), *The Psychology of Modern Prejudice*, Nova Science Publishers, pp. 123–146.

Lim, S.Y. and Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics*, 82(1): 145–164.

Lindberg, S.M., Hyde, J.S., Petersen, J.L., and Linn, M.C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, 136(6): 1123–1135.

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

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

McDonald, J. and Thornton, R. (2007). Do male and female college graduates receive unequal pay? *Journal of Human Resources*, 42(1): 32–48.

Moore, D.A. and Healy, P.J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2): 502–517.

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

Muzzatti, B. and Agnoli, F. (2007). Gender and mathematics: attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, 43(3): 747–759.

Niederle, M. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: the role of competition. *Journal of Economic Perspectives*, 24(2): 129–144.

Nix, S., Perez-Felkner, L., and Thomas, K. (2015) Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among STEM degree fields. *Frontiers in Psychology*, 6: 530.

O'Dea, R. E., Lagisz, M., Jennions, M. D., and Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, 9(1): 3777.

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

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

Perez-Felkner, L., Nix, S., and Thomas, K. (2017). Gendered pathways: how mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in Psychology*, 8: 386.

Plante, I., O'Keefe, P.A., Aronson, J., Fréchette-Simard, C., and 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. and Sydnor, J.R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24(2): 95–108.

Rapoport, B. and 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., and Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12): 44034408.

Reuben, E., Wiswall, M., and 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.

Ruissen, M.I. and de Bruijn, E.R. (2016). Competitive game play attenuates self-other integration during joint task performance. *Frontiers in Psychology*, 7: 274.

Steffens, M.C., Jelenec, P., and 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(4): 947–963.

Stoet, G. and Geary, D.C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4): 581–593.

Vuletich, H.A., Kurtz-Costes, B., Cooley, E., and Payne, B.K. (2020). Math and language gender stereotypes: age and gender differences in implicit biases and explicit beliefs. *PLoS One*, 15(9): e0238230.

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

A Experimental instructions [translated from Italian]

Thank you very much for taking part in this study, conducted by the University of Verona and the University of Trento, which involves eighth-grade students.

Your cooperation is very important. From now on we ask you not to communicate with the other participants. If you have any questions please raise your hand, and a research assistant will answer your questions individually.

The study consists of five parts. At the beginning of each part, you will receive detailed instructions for that part.

By participating in this study you will have the opportunity to receive a voucher that you can use to buy books at the Mondadori Bookstore located in S. Pietro Street in Trento. The voucher will be assigned to three randomly drawn participants in each class. The random draw will be made in class after the completion of PART 5.

The instructions for PART 1, PART 2, and PART 3 will detail how the total monetary value of the voucher is computed. Specifically, you can earn points that will be converted into vouchers at the conversion rate of 1 point = 1 euro if you are one of the winners in your class. The computation of the total amount of points will be made at the end of the whole study (approximately one month from now).

The answers you will provide will be anonymous. For this reason, we ask you to cut the ID tag you find at the bottom of this page and to keep it with you. If you are one of the winners, you will need this code to redeem you voucher.

The vouchers will be available at the bookstore at the end of the whole study.

Thank you again for your participation.

TO CUT AND KEEP

ID

□□□□□□

PART 1

In this part you will be asked to complete a quiz composed of 10 questions.

You will earn 1 point for each correct answer.

Some questions provide 4 answer options to select from (you have to select only one answer); other questions ask you to write the answer in the provided empty box. If you want to change an answer, cross it out and rewrite it.

You are not allowed to use calculators or other electronic devices. You can use the blank space on the quiz sheets for notes and calculations.

You have 15 minutes to complete the quiz. When we tell you that the time is up, you must stop writing, put the pen down on the table, and insert the quiz sheet into the provided transparent folder.

PART 2

[The gender presented first was counterbalanced across participants. We report the version with the female gender listed first.]

In this part you can earn additional points as explained below.

Several 8th grade students of middle schools in Trento take part in this study.

You will be asked to make two decisions, which are not linked to each other. One of the two decisions will be randomly drawn and the drawn decision will be relevant for the computation of your earned points.

DECISION 1

You must choose one of two randomly selected students participating in this study: a female 8th grader from a middle school different from yours (we call her "Participant A") and a male 8th grader from a middle school different from yours (we call him "Participant B"). Participants A and B are classmates.

[*Participants in the team-up treatment read:*

The choice you make will determine the points you will earn in this part. Specifically, you will earn 1 point for each correct answer that the participant you chose gave in PART 1 quiz.]

[*Participants in the competition treatment read:*

The choice you make will determine the points you will earn in this part. Specifically,

- if the number of correct answers you gave in PART 1 quiz is **greater than or equal to** the number of correct answers the participant you chose gave in the same quiz, you will earn 1 point for each correct answers you gave;
- if the number of correct answers you gave in PART 1 quiz is **lower than** the number of correct answers the participant you chose gave in the same quiz, you will not earn any points.]

To choose between Participant A and Participant B, you must divide a total of 10 virtual balls between A (the female student) and B (the male student). We will mark the balls you allocate to the female student with the letter "A" and those you allocate to the male student with the letter "B". The 10 marked balls will be put in a box and one of them will be drawn at random. The letter on the drawn ball will correspond to your choice and determine the points you will earn in this part.

Of course, the higher the number of balls allocated to A or B, the higher the probability that this participant will be drawn. Therefore, if you want one of the two participants to be drawn for sure, you should allocate all 10 balls to that participant; if you are indifferent between the participants, you should allocate 5 balls to each one of them; if you want one of the two participants to be drawn with a higher probability, you should allocate a higher number of balls to that participant. In any case, the sum of the balls you allocate to A and B must be equal to 10.

DECISION 2

You must once again choose one of two randomly selected students participating in this study. But now your choice is between a female or male 8th grader from your school (we call her/him "Participant A") and a female or male 8th grader from a middle school different from yours (we call her/him "Participant B").

[*Participants in the team-up treatment read:*

As for DECISION 1, you will earn 1 point for each correct answer that the participant you chose gave in PART 1 quiz. To choose between A and B, you must divide a total of 10 virtual balls, as in the procedure described for DECISION 1.]

[*Participants in the competition treatment read:*

The rules to compute the points you may earn and the dividing-balls procedure to choose between A and B are the same as those described for DECISION 1.]

The answer sheets used in Part 2 follow.

[Answers were collected separately for the two decisions.]

DECISION 1

You are asked to choose between Participant A and Participant B.
Participant A is a female 8th grader from a middle school in Trento different from yours.
Participant B is a male 8th grader from a middle school in Trento different from yours.
Participant A and Participant B are classmates.

YOUR CHOICE

To Participant **A**, a female 8th grader from a middle school in Trento	I choose to allocate no. ________ balls
To Participant **B**, a male 8th grader from a middle school in Trento	I choose to allocate no. ________ balls
	TOT 10

DECISION 2

You are asked to choose between Participant A and Participant B.
Participant A is a female or a male 8th grader from your school, but not your class.
Participant B is a female or a male 8th grader from a middle school in Trento different from yours.

YOUR CHOICE

To Participant **A**, a female or male 8th grader from my school, but not my class	I choose to allocate no. ________ balls
To Participant **B**, a female or male 8th grader from a middle school different from mine	I choose to allocate no. ________ balls
	TOT 10

PART 3

[The gender presented first was counterbalanced across participants and congruent with PART 2.]

In this part you can earn additional points. You will be asked to make six guesses and for each correct guess you will receive 1 point.

Q.1 How many correct answers do you think you gave in the ten-question quiz you completed in PART 1?

Write a whole number from 0 to 10 in this box: ☐

Q.2 How many correct answers do you think a female 8th grader from a middle school in Trento different from yours gave, on average, in the ten-question quiz (the same you completed in PART 1)?

Write a whole number from 0 to 10 in this box: ☐

Q.3 How many correct answers do you think a male 8th grader from a middle school in Trento different from yours gave, on average, in the ten-question quiz (the same you completed in PART 1)?

Write a whole number from 0 to 10 in this box: ☐

Q.4 How many correct answers do you think a female or male 8th grader from your school (but not your class) gave, on average, in the ten-question quiz (the same you completed in PART 1)?

Write a whole number from 0 to 10 in this box: ☐

Q.5 How many correct answers do you think a female or male 8th grader from a middle school in Trento different from yours gave, on average, in the ten-question quiz (the same you completed in PART 1)?

Write a whole number from 0 to 10 in this box: ☐

Q.6 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

PART 4

In this part you will receive two sheets. You will receive the second sheet after having handed in the first one.

Each sheet contains a list of words. The words belong to four different categories. Two of the categories are printed on the top left side of the list and the other two are printed on the top right side of the list. For each word, you are asked to indicate its appropriate category by making a check mark in the appropriate circle to the left or right of the word. You are given 45 seconds to categorise as many words as possible. At the end of 45 seconds, one of the assistants will announce that the time is up, and you will have to stop categorising and to put the pen down on the table.

You must categorise the words as fast and accurately as possible. Be sure to
- start from the top of the page and work down;
- avoid skipping any word;
- avoid correcting mistakes.

Below you can find an example with four categories (flower, insect, negative adjective, positive adjective) and some words (it is not necessary to solve the example):

Flower OR **Negative Adjective**		**Insect** OR **Positive Adjective**
◯	fun	◯
◯	cicada	◯
◯	unpleasant	◯
◯	ant	◯
◯	beautiful	◯
◯	roach	◯
◯	happy	◯
◯	rose	◯
◯	hard	◯
◯	cricket	◯
◯	likable	◯
◯	mimosa	◯
◯	annoying	◯
◯	ladybug	◯
◯	pleasant	◯
◯	cyclamen	◯
◯	noisy	◯
◯	iris	◯
◯	boring	◯
◯	daisy	◯

Stereotype-congruent sheet

[In the stereotype-incongruent sheet, "science" and "female" were paired on the right side.]

Science OR **Male**		**Liberal Arts** OR **Female**
○	Aunt	○
○	Biology	○
○	Grandfather	○
○	Chemistry	○
○	Father	○
○	Humanities	○
○	Female	○
○	Math	○
○	Woman	○
○	Astronomy	○
○	Husband	○
○	History	○
○	Wife	○
○	Geology	○
○	Grandmother	○
○	Literature	○
○	Mother	○
○	Engineering	○
○	Male	○
○	Philosophy	○
○	Uncle	○
○	Physics	○
○	Mother	○
○	Italian	○
○	Aunt	○
○	Math	○
○	Man	○
○	Arts	○
○	Woman	○
○	Biology	○
○	Wife	○
○	Music	○
○	Grandfather	○
○	Physics	○
○	Man	○
○	Humanities	○
○	Uncle	○
○	Italian	○
○	Father	○
○	Arts	○

PART 5

[We shall report only the survey questions relevant to our data analysis.]

Please read each question and mark the appropriate box or write your answer.

Q.1 You are male ☐ female ☐

Q.2 Which high school did you enrol in?

Type of high school	lyceum	☐	technical	☐	vocational	☐
Chosen curriculum						
Name of the school						

Q.3 As reported in the official high-school track recommendation you received, what type of high school and curriculum did your teachers recommended you to enrol in?

Type of high school ______________________ Curriculum ______________________

Q.4 Provide an answer for each of the following three questions:

		girls are definitely better	girls are somewhat better	there is no difference	boys are somewhat better	boys are definitely better
a)	In your opinion, who is better at math between girls and boys?	○	○	○	○	○
b)	According to your teachers, who is better at math between girls and boys?	○	○	○	○	○
c)	According to your classmates, who is better at math between girls and boys?	○	○	○	○	○

Q.5 Think about the math courses you have attended at school. How much do you think they have helped you

		not at all	to a small extent	to a moderate extent	to a large extent
a)	be aware of the applications of math in real life	○	○	○	○
b)	realise how math is relevant for everyday decisions	○	○	○	○
c)	get excited about math	○	○	○	○

Q.6 How much do you agree with the following statements about mathematics?

[The first nine statements compose the TIMSS (2015, 2019, grade 8) scale and the last five the Lim and Chapman (2013) subscale.]

		strongly disagree	moderately disagree	neither agree nor disagree	moderately agree	strongly agree
a)	I usually do well in math	○	○	○	○	○
b)	Math is more difficult for me than for many of my classmates	○	○	○	○	○
c)	Math is not one of my strengths	○	○	○	○	○
d)	I learn things quickly in math	○	○	○	○	○
e)	Math makes me nervous	○	○	○	○	○
f)	I am good at working out difficult math problems	○	○	○	○	○
g)	My teacher tells me I am good at math	○	○	○	○	○
h)	Math is harder for me than any other subject	○	○	○	○	○
i)	Math makes me confused	○	○	○	○	○
j)	Math is a very worthwhile and necessary subject	○	○	○	○	○
k)	Math is important in everyday life	○	○	○	○	○
l)	Math is one of the most important subjects for people to study	○	○	○	○	○
m)	Math lessons are very helpful no matter what I decide to study in future	○	○	○	○	○
n)	A strong math background could help me in my professional life	○	○	○	○	○

B Description of regression variables (see Tables 3 and 5)

Variable	Min	Med	Max	Mean	SD	N
Girls						
STEM	0.000	0.000	1.000	0.282	0.451	248
MATH.Q	1.000	2.000	4.000	2.319	1.045	248
MATH.belief	0.000	7.000	10.000	6.427	1.823	248
MATH.median	0.000	0.000	1.000	0.268	0.444	231
Matching decision	-10.000	0.000	10.000	-0.129	4.558	248
Diff Exp Performance	-8.000	0.000	9.000	-0.423	1.517	248
Own implicit bias	-8.375	1.999	20.785	1.474	4.382	240
Explicit bias index	-2.000	0.000	2.000	-0.119	0.931	240
Boys						
STEM	0.000	0.000	1.000	0.472	0.500	286
MATH.Q	1.000	2.000	4.000	2.395	1.152	286
MATH.belief	0.000	7.000	10.000	6.993	2.003	286
MATH.median	0.000	1.000	1.000	0.504	0.501	262
Matching decision	-10.000	0.000	10.000	-0.671	4.936	286
Diff Exp Performance	-10.000	-1.000	7.000	-0.472	1.754	286
Own implicit bias	-13.297	-1.042	17.500	0.172	4.620	273
Explicit bias index	-2.000	0.000	2.000	0.157	1.115	272