

# Implicit Stereotypes: Evidence from Teachers' Gender Bias\*

Michela Carlana<sup>†</sup>

January 2019

## Abstract

I study whether exposure to teacher stereotypes, as measured by the *Gender-Science Implicit Association Test*, affects student achievement. I provide evidence that the gender gap in math performance, defined as the score of boys minus the score of girls in standardized tests, substantially increases when students are assigned to math teachers with stronger gender stereotypes. Teacher stereotypes induce girls to under-perform in math and self-select into less demanding high-schools, following the track recommendation of their teachers. These effects are at least partially driven by a lower self-confidence on own math ability of girls exposed to gender biased teachers. Stereotypes impair the test performance of girls, who end up failing to achieve their full potential. I do not detect statistically significant effects on student outcomes of literature teacher stereotypes.

**JEL:** J16, J24, I24.

**Keywords:** gender, math performance, teachers, implicit stereotypes, IAT, self-confidence, track choice.

---

\*I am grateful to four anonymous referees, Eliana La Ferrara, Nicola Gennaioli, Jennifer Lerner, Valerio Nispi Landi, Dan-Olof Rooth, David Stromberg, Diego Ubfal, and seminar participants at the NBER Education SI, Harvard Kennedy School, Boston University, University of Zurich, Universitat Pompeu Fabra, University of Maryland (School of Public Policy), McGill University, World Bank, FAIR- Norwegian School of Economics, IIES, Stockholm University, University of Bologna, University of Vienna, Central European University, IZA, COSME 2018 Gender Workshop, ECBE 2018 Bergen, OxDev 2017, 32nd AIEL Conference, and SSE Human Capital Workshop 2017 for their extremely helpful comments. Elena De Gioannis and Giulia Tomaselli provided invaluable help with data collection. This paper is funded under the grant "Policy Design and Evaluation Research in Developing Countries" Initial Training Network (PODER), which is financed under the Marie Curie Actions of the EU's Seventh Framework Programme (Contract Number: 608109) and received financial support from the Laboratory for Effective Anti-poverty Policies (LEAP-Bocconi). I am indebted to Gianna Barbieri and Lucia De Fabrizio (Italian Ministry of Education, Statistics), Patrizia Falzetti and Paola Giangiacomo (Invalsi) for generous support in providing the data. I am grateful to all principals and teachers of schools involved in this research for their collaboration in data collection. I thank Pamela Campa for providing World Value Survey data on Italian provinces. This research project was approved by the Ethics Committee of Bocconi University on the 14th September 2016.

<sup>†</sup>Harvard Kennedy School, 79 John F. Kennedy, Cambridge (MA) (e-mail: michela\_carlana@hks.harvard.edu).

# 1 Introduction

Over the last century, the narrowing of gender differences in labor market participation and educational outcomes has been impressive, even reversing the gap in school attainment (Goldin et al., 2006). In spite of this, boys outperform girls in math, with an even wider gap among the highest-achieving students and with potential consequences for the under-representation of women in highly profitable fields (OECD, 2014). Math performance has been shown to be a good predictor of readiness for science, technology, engineering, and math (STEM) universities and future labor market outcomes (Altonji and Blank, 1999; Card and Payne, 2017). There is a long-standing debate on whether the gender gap in math achievement arise from biologically based differences in brain functioning as opposed to culture and social conditioning (Baron-Cohen, 2003; Nollenberger et al., 2016). Cross-country evidence supports the latter idea: cultures in which gender stereotypes are weaker have a smaller gender gap in math performance, defined as the score of boys minus the score of girls in standardized tests (Guiso et al., 2008; Nosek et al., 2009; Else-Quest et al., 2010).<sup>1</sup>

Stereotypes may induce discrimination if own preconceived beliefs interfere with the ability to be impartial or if they impair group members' performance (Glover et al., 2017; Bohren et al., 2018).<sup>2</sup> Without provision of further information about the candidates except their appearance, men are more likely to be hired for a mathematical task than women (Reuben et al., 2014), and both men and women are less willing to contribute ideas and have lower self-confidence in fields that are not stereotypically associated with their own gender (Coffman, 2014; Bordalo et al., 2018). However, whether exposure to gender stereotypes in the real world affects the emergence of the gap in math and reading skills remains an empirical question.

Stereotypes communicated by teachers may be particularly detrimental for children, as they

---

<sup>1</sup>For instance, Nosek et al. (2009) exploits the Gender-Science Implicit Association Test to measure stereotypes and finds that it predicts nation-level sex differences in 8th-grade science and mathematics achievement.

<sup>2</sup>Stereotypes are mental constructs based on over-generalized representations of differences between groups (Bordalo et al., 2016). I define discrimination following Bertrand and Duflo (2017): "members of a minority group (...) are treated differentially (less favorably) than members of a majority group with otherwise identical characteristics in similar circumstances".

affect the development of academic self-concept (Ertl et al., 2017). According to research in social psychology, teachers are likely to believe math is more difficult for girls than for equally achieving boys (Riegle-Crumb and Humphries, 2012; Tiedemann, 2002) and they implicitly convey their stereotyping through their classroom instruction (Keller, 2001). Teachers' erroneous expectations may lead to a self-fulfilling prophecy whereby prior beliefs are self-confirming in equilibrium (Spencer et al., 1999; Papageorge et al., 2018): biased teachers may set a lower bar for the learning of students from stigmatized groups or fail to encourage them to fulfil their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983).

This article documents the impact of exposure to teacher stereotypes during middle school on student outcomes, including standardized test scores in math and reading, choice of the field of study, and self-confidence.

One of the main challenges to address this question is the availability of an appropriate measure of teacher stereotypes matched with students' achievements and choices. I focus on the Italian context and build a unique dataset including administrative information and surveys. I measure stereotypes of around 1,400 math and literature teachers working in 102 schools in the North of Italy using the Gender-Science Implicit Association Test (IAT). This test is a computer-based tool developed by social psychologists (Greenwald et al., 1998) and recently used by economists when studying discrimination in the context of gender and race bias (Rooth, 2010; Corno et al., 2018; Glover et al., 2017). The test exploits the reaction time to associations between male or female names and scientific or humanistic fields. The underlying assumption is that responses are faster and more accurate when gender and field subjects are more closely associated by the individual (Lane et al., 2007).

I document that implicit associations, measured by the IAT, reflect stereotypes based on the representativeness of genders at the top of the ability distribution for math and reading (Bordalo et al., 2016). In addition to IAT scores, I collected detailed information on teacher characteristics. I show that IAT scores correlate with observables, including gender, field of study, and gender norms in the place of birth, as measured by the World Value Survey. Furthermore, I find

that IAT scores do not correlate with variables such as teachers' experience, or self-reported gender bias, which could arise either because they measure two different mental constructs or because there is social desirability bias in the explicit answers (Greenwald et al., 2009).

I link the teacher survey with administrative information on pupils from the Italian Ministry of Education and the National Institute for the Evaluation of the Italian Education System (INVALSI), and with a newly collected student questionnaire. Data on pupils include standardized test scores in math and reading, family background, high-school track choice, teachers' track recommendation and, for a sub-sample of students, a measure of self-confidence in own abilities in different subjects.

The identification strategy relies on the "as good as random" assignment of students to teachers with different levels of implicit stereotypes. I provide supporting evidence showing that baseline characteristics of students, such as family background, are not systematically correlated with teacher stereotypes. I use two identification strategies. First, I focus on gender gaps within classes, including class fixed effects that absorb all characteristics of peers, the school environment, and teachers. I exploit variation in performance and track choice between boys and girls enrolled in the same class.<sup>3</sup> Second, I compare students of the same gender, enrolled in the same school and cohort, but assigned to teachers with different levels of stereotyping. This exercise explores whether the wider gender gap in classes assigned to teachers with stronger stereotypes is due to girls lagging behind, boys improving more, or a combination of the two effects.

I find that math teachers with stronger implicit stereotypes, as measured by the Gender-Science IAT, have a negative and quantitatively significant influence on girls. The gender gap in math performance in grade 8 increases by 0.03 standard deviations when students are assigned to teachers with one standard deviation higher implicit stereotype score during middle school. In other words, the gender gap in math performance increases by one third (from 0.15 to 0.20 standard deviations) in classes assigned to a math teacher who implicitly associates boys with

---

<sup>3</sup>Students are assigned to the same group of peers from grade 6 to grade 8. Teachers are assigned to classes and follow students during all years of middle school, with few exceptions due to retirement or transfers.

mathematics compared to classes assigned to a teacher who has the opposite implicit associations. Teacher stereotypes have no effect on boys, while they lower math scores for girls, especially those with lower initial performance.

Stereotypes of literature teachers have no effect on reading performance of boys or of girls. Several reasons may explain the asymmetric impacts of stereotypes by subjects. Girls may be more vulnerable to the gender stereotype that “*women are bad at math*” than boys are to the gender stereotype that “*men are bad at reading*”, consistent with [Kugler et al. \(2017\)](#) and [Große and Riener \(2010\)](#), or teachers may be more likely to convey their stereotyping through their classroom instruction in math than in literature ([Keller, 2001](#)).

Next, using an ordered logit with class fixed effects (BUC estimator, following [Baetschmann et al., 2015](#)), I provide evidence that math teacher stereotypes induce girls to self-select into less demanding tracks, following the biased recommendation of their teachers. The estimates from a linear probability model suggest that a substantial part of the effect is driven by a higher likelihood to enrol in the vocational track for girls exposed to teachers with stronger implicit stereotypes. The effect is driven by students at the bottom of the ability distribution or with missing data on test scores.<sup>4</sup> These results provide a link between teacher stereotypes and teacher bias: they suggest that stronger male-math implicit associations of teachers interfere with their interaction with female students and their ability to be unbiased in the classroom, even unconsciously – for instance, when they recommend a high-school track to their students.

Finally, I show that teacher stereotypes have a substantial negative impact on girls’ self-confidence in math. The finding is consistent with the hypothesis that stereotypes impair the test performance of ability-stigmatized groups, who end up failing to achieve their full potential. This is a crucial channel to explain the underperformance of girls in math when assigned to more biased teachers, but is also broadly relevant since it suggests that the lower self-confidence of women in the scientific fields is at least partially activated by exposure to gender stereotypes.

---

<sup>4</sup>For the sample of student without missing data on test scores, the impact is smaller and indistinguishable from zero. Data on test scores are missing if the student did not take the test or if the school did not provide the correct match between administrative data from the Ministry of Education and INVALSI.

Implicit stereotypes create a self-fulfilling prophecy, perpetuating gender differences in math performance.

This study adds to the recent literature in economics that has uncovered the benefit of incorporating insights from social psychology and considering *implicit bias* in studying discrimination (Guryan and Charles, 2013; Bertrand and Duflo, 2017). My article investigates the role of *implicit* stereotypes in the context of education economics and pupil-teacher interactions. Implicit stereotypes can operate even without awareness or intention to harm the stigmatized group (Bertrand et al., 2005). In particular, we may expect that teachers do not explicitly endorse gender stereotypes, but their implicit stereotypes, embedded in their own experiences since childhood, affect their interaction with pupils. My work also contributes to the debate in the social psychological literature on what the IAT is measuring and on its predictive power of actual behaviour (McConnell and Leibold, 2001; Blanton et al., 2009; Oswald et al., 2013).

Teachers matter for students' performance and later-life outcomes (Chetty et al., 2014a,b) and their gender stereotypes may be an important channel. The economics literature analyzing the impact of teacher gender stereotypes on student outcomes has mainly focused on either self-reported measures (Alan et al., 2018) or bias in grading, i.e. the gender differences in grades given in blind vs. open evaluations (Lavy and Sand, 2018; Lavy and Megalokonomou, 2017). Compared to other measures of teacher bias, the IAT has two main advantages. First, it does not suffer from social desirability bias that may be an issue in self-reported measures. Second, stereotypes are measured without relying on student performance, which may capture variation in unobservable characteristics of pupils potentially correlated with future outcomes.

A growing number of papers exploits the gender of teachers as a proxy of their pupils' exposure to stereotypes and role-modeling (Bettinger and Long, 2005; Dee, 2005; Carrell et al., 2010; Antecol et al., 2014). In this paper, I provide evidence that the gender of teachers is correlated with the Gender-Science IAT score and that the impact of implicit stereotypes on student outcomes is slightly stronger for male teachers, compared to female.

Finally, I contribute to understanding the importance of gender-biased environments in ex-

plaining the under-confidence of females in STEM fields. Gender differences in confidence and competitiveness have negative consequences for women’s performance, as well as their educational and occupational choices (Coffman, 2014; Buser et al., 2014; Reuben et al., 2015; Kugler et al., 2017). Exposure to biased teachers activates negative self-stereotypes in female students. The results are consistent with predictions of the stereotype threat theory, according to which individuals at risk of confirming widely-known negative stereotypes reduce their confidence and underperform in fields in which their group is ability-stigmatized (Spencer et al., 1999).

## 2 Setting

In the Italian educational system, middle school lasts three years, from age 12 to 14. Students in middle school are assigned to classes at the beginning of grade 6 and stay with the same peers until the end of grade 8.<sup>5</sup> The general class formation criteria are established by an Italian law and details are specified by each school council in a formal document available on the website of the institution. The general criteria mentioned by most schools are equal allocation of students across classes according to gender, disability, socioeconomic status, and ability level (as reported by the elementary school).<sup>6</sup> Moreover, I collect information directly from principals on how classes are formed. School principals report that the most relevant aspects in the class formation process are comparability across classes and heterogeneity within classes in the same school (for detailed information, see Online Appendix B). What is important for my analysis is that I can test whether this intention of the principals is confirmed by the allocation of students to classes in my sample (see section 4.3).

Teachers are assigned to schools by the Italian Ministry of Education, and their salary is

---

<sup>5</sup>There are a few exceptions: students may be transferred to a different school by their parents or be required by their teachers to repeat a grade.

<sup>6</sup>The D.P.R. 20 marzo 2009 n.81 establishes, for instance, that the number of students per class in middle school should be between 18 and 27. Further information at school level is provided on the “Plan of Education Offer” (“Piano dell’Offerta Formativa”). An analysis of Ferrer-Esteban (2011) shows that ability grouping across classes within schools occurs almost exclusively in the South of Italy. All schools in my sample are from the North of Italy.

determined by experience in a centralized system. Teachers' allocation across schools is settled by seniority: when teachers accumulate years of experience, they tend to move closer to their home town and away from disadvantaged areas (Barbieri et al., 2011). Each class is assigned by the principal to a math and literature teacher among those available in the school, and teachers usually follow students from grade 6 to grade 8. Every week, students spend at least 6 hours with the math teacher and 5 hours with the literature teacher.<sup>7</sup>

Standardized tests in math and reading are administered in grades 2, 5, 6, 8 and 10 by the National Institute for the Evaluation of the Italian Education System (INVALSI).<sup>8</sup> The tests are presented to all students as ability tests, thus making the gender stereotype in math potentially relevant. They are graded anonymously following a precise evaluation grid and by a different teacher than the one instructing students in the specific subject. Students are not informed about their performance on the test, except in grade 8. The grade 8 achievement test score has higher stakes since, until 2017, it affected 1/6 of the final score of students at the end of middle school. Test scores have no direct impact on enrolment in high school, but they are highly predictive of students' high-school track choice (Carlana et al., 2019) and potentially of their later labor market outcomes, as shown in other countries (Murnane et al., 1995; Meghir and Palme, 2005).<sup>9</sup>

After middle school, students self-select into three different tracks: academic oriented ("liceo"), technical, and vocational high school. Each type of school is divided into several subtracks: the academic oriented track can be specialized in either scientific, humanistic, languages, human sciences, artistic or musical subjects. The technical track can be focused on technological or economic subjects; and the vocational track can have different core subjects, for instance hospitality training, cosmetics, and mechanical workshops. Students are free to choose a high school

---

<sup>7</sup>Students can be enrolled in school from 30 to 43 hours per week and, therefore, the amount of time they spend with teachers varies. For instance, they spend from 6 to 9 hours with the math teacher. In some classes, literature teachers also teach history and geography so they spend more time with students. The number of hours per week spent with the literature teacher, therefore, varies from 5 to 10.

<sup>8</sup>The test in grade 6 was administered only up to the school year 2012-13. All students are supposed to take the test, unless they are absent from school on the day of the test. It may also happen that the school misreports the code that allows to match the test score with the administrative data from the Ministry of Education. This happened for 7% of cases in grade 8 for the sample of schools used in this paper.

<sup>9</sup>In Italy, standardized test score data have never been matched with labor market outcomes.

with no restriction on the track based on grades or ability, and they tend to choose according to family background and the child's enjoyment of the curriculum (Giustinelli, 2016). Teachers give a non-binding track recommendation to families with an official letter sent to each child's home, which is also reported to the Ministry of Education.

The choice of high-school is strongly correlated with university choice: 80% of graduates in STEM universities in 2015 did a scientific academic or a technical track during high-school (62% did the scientific academic high-school track). Among students enrolled in the vocational track, only 1.7% of the cohort graduating in 2016 enrolled in university, while the percentage increased to 73.7% and 32.3% in the academic and technical tracks respectively. Interestingly, the majority of technical track students enrol in either STEM or economics degrees: 62.5% vs. 52.4% of the academic track students.

## 3 Data

### 3.1 Sample

During September 2016, I invited 145 middle schools to take part in a research project regarding "The role of teachers in high-school track choice," out of which 102 accepted and 91 provided all information necessary for my study.<sup>10</sup> The sample was designed to include all schools of the provinces of Milan, Brescia, Padua, Genoa and Turin with more than 20 immigrants enrolled in grade 6 in the school year 2011-12.<sup>11</sup> The Appendix Table A.I shows the balance tables of the characteristics of students used in the analysis and those of all Italian students in the same

---

<sup>10</sup>In 102 schools, I obtained the authorization of the principal to administer the survey to teachers, but only 91 principals completed (without mistakes) the formal authorization to give me access to data from the National Institute for the Evaluation of the Italian Education System (INVALSI).

<sup>11</sup>The data collection was also conducted for ongoing work studying teacher race stereotypes (Alesina et al., 2018).

cohorts.<sup>12</sup> Although the standardized difference is always below the cutoff of 0.25 suggested by Imbens and Rubin (2015), as expected, the sample used in this paper has a higher share of immigrants compared to the national average (21.7% vs. 9.6%) and also compared to the average of the five provinces in the North (21.7% vs. 13.4%).<sup>13</sup> The average math scores of both boys and girls are similar to the local and national average.

### 3.2 Data Sources

From October 2016 to March 2017, I conducted a survey of around 1,400 math and literature teachers. The questionnaire was administered directly by enumerators using tablets in a meeting held in school buildings. Participants agreed to take part in the survey and signed an informed consent, in which it was explained that the survey was part of a research project aimed at analyzing the role of teachers in affecting students' track choices. There was no reference to gender bias. The time to complete the survey was around 30 minutes and teachers did not receive compensation for taking it. Among all math and literature teachers working in the schools involved in this research, around 80 percent completed our survey thanks to the strong support of principals.<sup>14</sup> The survey was divided into two parts: the Implicit Association Test (IAT) and a questionnaire.

On top of the teacher survey data, I use three additional sources of data: student survey data, administrative information from the Italian Ministry of Education (MIUR) and from the

---

<sup>12</sup>The normalized difference shown in column 4 is the formula recommended by Imbens and Wooldridge (2009):

$$\Delta = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2}}$$

where  $\bar{X}_1$  and  $\bar{X}_2$  are the means of covariate  $X$  in the two sub-groups that are being compared, and  $S_1^2$  and  $S_2^2$  are the corresponding sample variances of  $X$ . Imbens and Rubin (2015) recommend, as a rule of thumb, that  $\Delta$  should not exceed 0.25.

<sup>13</sup>Around half the students are first generation and half are second generation immigrants.

<sup>14</sup>Only 4 math teachers started the questionnaire and then did not finish it since they claimed either that they were not expecting such a long survey or that they could not understand the purpose of the Implicit Association Test. I prepared a report for each principal at schools where more than 70% of teachers completed the survey with summary statistics on the outcomes of their students during high-school to encourage principals to increase teacher participation. The report was delivered to schools during the summer of 2017, after the middle school graduation of the 2017 cohort.

National Center for the Evaluation of the Italian Educational System (INVALSI).

### ***A1. Implicit Association Test***

I measure implicit gender stereotypes using the IAT, a tool developed within social psychology (Greenwald et al., 1998; Lane et al., 2007). The IAT uses the categorization of words to the left or to the right of a computer or tablet screen to provide a measurement of the strength of the association between two concepts – in this case, gender and scientific/humanistic fields. Subjects were presented with two sets of stimuli. The first set included names of females (e.g., Anna) and males (e.g., Luca), and the second set included subjects related to scientific (e.g., Calculus) and humanistic fields (e.g., Literature). Words appear one at a time at the center of the screen, and individuals are instructed to categorize them as fast as possible to the left or the right according to different labels displayed on the top of the screen (for instance, on the right the label “Female” and on the left the label “Male”). To calculate the score, two types of tasks are used: in the first task, individuals are instructed to categorize male names and scientific subjects to one side of the screen and, on the opposite side of the screen, categorize female names and humanistic subjects (“order compatible” task), while in the second task, individuals are instructed to categorize to one side of the screen female names and scientific subjects and to the opposite side of the screen male names and humanistic subjects (“order incompatible” task). The idea behind the IAT is that if individuals have implicit associations between men and scientific fields, it should be easier and quicker to do the task when they categorize these words on the same side of the screen. A detailed explanation of the IAT is provided in the Online Appendix C.<sup>15</sup>

A broad strand of literature in social psychology, and an increasing number of papers in economics, has provided evidence on the validity of IAT scores in predicting relevant choices and behaviors (Nosek et al., 2007; Greenwald et al., 2009). For example, Reuben et al. (2014) shows

---

<sup>15</sup>The order of the tasks was randomized at individual level and in the Online Table C.I I provide evidence that the impact of the order of the blocks is small in magnitude. However, in all regressions, I control for ordering factors, but they do not have a statistically or economically significant effect on the estimates.

in a lab experiment that higher stereotypes (measured by the Gender-Science IAT) predict employers' biased expectations against female math performance and also predict the suboptimal update of expectations after ability is revealed. Higher implicit gender bias is acquired at the beginning of elementary school and is generally associated with lower performance of females in math during college, lower desire to pursue STEM-based careers, and lower association of math with themselves, even for women who select math-intensive majors (Cvencek et al., 2011; Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007).<sup>16</sup>

There is a lively debate among social psychologists on Implicit Association Tests. First, some papers have argued that the IAT has weak predictive validity (Blanton et al., 2009; Oswald et al., 2013). Most of the studies refer to experiments with fewer than 50 subjects and do not have information outside the lab on whether individuals with stronger implicit associations are actually biased in their interaction with stigmatized groups. I believe that further research is necessary and my paper can contribute to this debate. Second, some studies suggest that IAT results can be faked after respondents acquire knowledge of the test (Fiedler and Bluemke, 2005). The IAT is not widespread in Italy, and none of the teachers who took the survey reported familiarity with the test. Without any hints, it seems unlikely that they were able to figure out how to trick the test. Third, IAT scores, at least partially, capture unstable characteristics that do vary over time.<sup>17</sup> This short-term exposure may introduce additional noise in the measurement, leading to an attenuation bias when I estimate the impact of teacher stereotypes on student outcomes. Finally, IAT scores could be contaminated by *extrapersonal associations* that are available in memory, but do not contribute to an individual's personal evaluation when one interacts with the specific category (Olson and Fazio, 2004) or they may reflect "cultural stereotypes rather than personal animus" (Arkes and Tetlock, 2004). The concern of capturing associations outside the schooling context is alleviated given that teachers complete the survey within the school building by associating school subjects with gender. However, as I document in Section 4.2, there is

---

<sup>16</sup>In the context of race implicit bias, studies have shown the relevance of IAT scores in affecting job performance of minorities (Glover et al., 2017) and call-back rates of job applicants (Rooth, 2010).

<sup>17</sup>For instance, race implicit associations have been shown to decrease after subjects viewed pictures of admired African Americans and disliked White Americans (Dasgupta and Greenwald, 2001).

a significant correlation between Gender-Science IAT scores and gender norms in the place of birth of individuals. In my opinion, this fact does not undermine the importance of studying the impact of implicit stereotypes.

To sum up, IAT scores are a noisy measure of implicit stereotypes that may be affected by culture and socialization. Nevertheless, they have the great advantage of avoiding social desirability bias in the response and capturing implicit associations potentially unconscious to the individual that may affect his or her interaction with the stigmatized group. In this study, I am not interested in whether teachers have stereotypes (i.e. in the level of IAT score), but on whether those with higher stereotypes have a negative impact on student outcomes.

## ***A2. Teachers' Questionnaire***

After the Implicit Association Tests, enumerators invited teachers to complete a questionnaire asking detailed information about family background of teachers (age, parents' education, place of birth, age and sex of children, etc.) and career related aspects (type of contract, years of experience, whether they are involved in the management of the school or in the organization of Math Olympics Games, refresher courses, etc.). Furthermore, they were also asked questions about explicit bias, for instance beliefs about gender differences in innate math ability and the standard Word Value Survey question: "*When jobs are scarce, men should have more right to a job than women*".<sup>18</sup> Participants are in general reluctant to explicitly endorse gender stereotypes (Nosek et al., 2002), potentially leading to social desirability bias in the responses. These aspects are emphasized by the awareness of being interviewed as teachers.

Enumerators collected data on the allocation of teachers to classes from the school year 2011-12 to the school year 2016-17, in order to merge teacher and student data. I double check all this information using data provided directly by schools and on their websites.

---

<sup>18</sup>The specific questions are reported in the Online Appendix [C.2](#).

## ***B. Administrative Data and Students' Self-Confidence***

I obtained student-level information from the Italian Ministry of Education and from the National Institute for the Evaluation of the Italian Education System (INVALSI) for the cohort of students enrolled in grade 8 between school year 2011-12 and 2016-17.<sup>19</sup> The data available include math and reading standardized test scores in grade 8 and grade 6<sup>20</sup>, parents' education and occupation, baseline individual information (date and place of birth, gender, citizenship), high-school track choice, and official teachers' recommendation. In 2014, students in grade 8 at 24 schools in this sample were asked to complete a survey about their track choice around two months before the end of the school year. In particular, they reported their belief about their own ability in each subject, choosing between “good”, “mediocre”, and “scarce”.<sup>21</sup>

### **3.3 Descriptive Statistics**

#### ***Teachers***

The dataset includes 537 math and 853 literature teachers, but I restricted the main analysis to 454 math and 615 literature teachers (“matched sample”) for whom I have student data for grade 8.<sup>22</sup> Table I reports descriptive statistics for several teachers' characteristics. Most teachers are females (81% for math and 90% for literature). They are on average 49 years old with 20 years of experience in teaching, and 79% of math teachers and 94% of literature teachers hold a full-time contract. The majority (61% for math and 72% for literature) of teachers were born in a

---

<sup>19</sup>Individual level data are anonymous and I obtained the authorization from each school principal to access data from their school. The data from the Italian Ministry of Education is available only up to the school year 2015-16.

<sup>20</sup>The standardized test score in grade 6 is available only up to 2012-13. The test was not administered anymore after that year.

<sup>21</sup>The specific question is reported in the Online Appendix C.3.

<sup>22</sup>As discussed above, 11 principals did not complete (without mistakes) the formal authorization to give me access to all data. Furthermore, I have to exclude teachers who did not teach in grade 8 and for whom I do not have student outcomes. Finally, 3 math teachers and 9 literature teachers did not complete the Gender-Science IAT test. Online Appendix Table A.II shows the balance table of the differences between the sample of teachers matched and the other teachers who completed the IAT. As expected, teachers not matched are around 9 years younger, and 35 percent less likely to have a full-time contract (tenured position), and they have 11 years fewer of experience in teaching. However, not only the average, but also the entire distribution of implicit gender bias of the matched and not-matched teachers is extremely close (exact p-value of Kolmogorov-Smirnov: 0.590 for math teachers and 0.466 for literature teachers, Online Appendix Figure A.I).

city in the North of Italy, but a substantial share were born in the Center or South of Italy and then migrated to the North to work. Most math teachers graduated from programs in biology, natural sciences and other related subjects: only 23% studied math, physics, or engineering.

Considering the IAT thresholds typically used in the social psychological literature, 16% of teachers associate math with girls, 23% present little to no clear association, 19% show male-math association and 42% show moderate to severe male-math association.<sup>23</sup> For comparison, the sample of 1164 Italians used by Nosek et al. (2009) has an average Gender-Science IAT score of 0.40 (SD 0.40): the score of math teachers is on average substantially lower (mean 0.09, SD 0.37, as shown in Table I), while literature teachers are very close to this mean (mean 0.38, SD 0.39).<sup>24</sup> Notably, the great majority of math teachers are women and this may have important implications for the association of scientific subjects with gender (see further discussion in Section 4.2). For ease of interpretation of my results, I standardize the IAT score to have mean zero and variance one in the main results of the paper.

At the bottom of Table I, I report the summary statistics of explicit stereotypes described in detail in the Online Appendix C. There is little variability in the self-reported bias questions, potentially also due to social desirability bias and the widespread explicit rejection of stereotypes. Teacher “quality” is proxied with factors that are usually positively evaluated by teachers and principals: years of experience in teaching, full-time contract, and being the teacher in charge of math Olympics in the school. Online Appendix Table A.III shows that these factors are correlated with student performance on standardized tests in math. The results for literature go in the same direction, but are smaller and statistically indistinguishable from zero. The effect of teacher quality on student performance is similar for girls and boys.

---

<sup>23</sup>Greenwald et al. (2003) suggests that a raw IAT score below -0.15 shows bias in favour of the stigmatized group, between -0.15 and 0.15 little to no bias, from 0.15 to 0.35 slight bias against the stigmatized group and a value higher than 0.35 as moderate to severe bias against the stigmatized group. The distribution of IAT score is plotted in Figure I.

<sup>24</sup>In the paper by Nosek et al. (2009), individuals completed the IAT online at the *Implicit Project* website.

## *Students*

Table II reports summary statistics on students. I restrict the sample to students with a standardized test score in grade 8 and for whom I have the implicit association test of their math teacher in grade 8.<sup>25</sup> In the sample, 51.7% of students are males and boys and girls are balanced in terms of baseline characteristics related to place of birth, generation of immigration, and parents' education and occupation. Test scores are standardized by subject and year to have mean zero and standard deviation one. Females at the beginning of middle school score 0.19 standard deviations lower in math and 0.14 standard deviations higher in reading than males do. In the same table, I also report the raw and normalized gender differences in outcomes (Imbens and Wooldridge, 2009).

High-school track choices in this sample are comparable to the national average: females are 10 percentage points less likely to choose an academic scientific track and almost 25 percentage points less likely to enroll in a technical technological track. Girls are more likely to choose an academic track than boys, but not top-tier ones, which include classical and scientific tracks. Vocational school is chosen at an equal rate by both genders. However, teachers recommend 38.5% of males to the vocational track and 31.5% of females, while the scientific track is recommended only to 18% of males and 13% of females.<sup>26</sup>

Using the student survey data, I document that, on average, there are no gender differences in assessment of ability, but females are 8 percentage points less likely than boys to consider themselves good at math and boys are 5 percentage points less likely to consider themselves good at reading.

---

<sup>25</sup>I only consider classes with at least ten students with standardized test scores.

<sup>26</sup>In some schools, more than one recommendation is given to students. I consider whether at least one of the choices recommended was scientific or vocational. The results are substantively identical when I consider only the first choice of teachers' recommendation.

## 4 Empirical Strategy

### 4.1 Estimating Equation

The main purpose of this paper is to investigate the impact of teacher stereotypes on student achievement. I exploit two identification strategies. The first is aimed at investigating the gender gap within a class, estimating the following equation:

$$y_{ic} = \alpha_0 + \alpha_1(Female_i \times stereotypes_c) + \alpha_2 Female_i + \eta_c + \mathbf{X}_i \rho_1 + (Female_i \times \mathbf{X}_i) \rho_2 + (Female_i \times \mathbf{Z}_c) \rho_3 + \varepsilon_{ic} \quad (1)$$

where  $y_{ic}$  is the outcome (i.e., math standardized test score, track choice, and self-confidence) of student  $i$  in class  $c$ .  $Female_i$  is a dummy variable which assumes value 1 if the student  $i$  is a girl and  $stereotypes_c$  is the standardized value of the IAT score of the math teacher assigned to class  $c$  in grade 8.<sup>27</sup> I include fixed effects at the class level  $\eta_c$ , which absorb the average effect of teacher bias in class  $c$ . Furthermore, I include student characteristics  $\mathbf{X}_i$  (parental education and occupation, immigration status and generation of immigration), and teacher characteristics  $\mathbf{Z}_c$  (gender, place of birth, age, teacher “quality”, type of contract, and type of degree achieved) interacted with the gender of student  $i$ . Standard errors are robust and clustered at the teacher level.

Crucially, in this identification strategy, class, teacher, and school level characteristics are absorbed by class fixed effects. Indeed, as described in Section 2, students are assigned to a class in grade 6 and attend all classes with the same classmates until grade 8. We can only identify the impact of teachers’ IAT scores on the gender gap in the dependent variable, i.e. the interaction between the gender of students and implicit stereotypes of teachers. The coefficient of interest,  $\alpha_1$ , measures how the gender gap in the class is affected by the assignment to teachers with

---

<sup>27</sup>In 63% of the cases math teachers have been teaching to the same class from grade 6 to grade 8, in 14% of the cases from grade 7 and in 22% only for grade 8. This information includes only the cohorts who began grade 6 from 2011-12 and for which I collected information on the teacher assignment for all three years of middle school. Two or three different classes can be assigned to the same teacher.

one standard deviation higher stereotypes.<sup>28</sup> I expect the estimate of  $\alpha_1$  to be attenuated as a result of the measurement error in the gender IAT score. Indeed, occasion-specific noise may introduce an attenuation bias, as suggested by Glover et al. (2017).<sup>29</sup> For robustness, I include controls for student characteristics  $\mathbf{X}_i$  interacted with the gender of the pupil. The regression also controls for the gender of students interacted with teacher characteristics  $\mathbf{Z}_c$ . This is potentially important to partial out the differential impacts on boys and girls by gender, background, and other observable characteristics of teachers. Furthermore, this allows me to establish whether the impact of teacher stereotypes on the gender gap within the class can be explained (or attenuated) by teachers' observables.

The second identification strategy relies on the comparison of students of the same gender enrolled in the same school, but assigned to teachers with different stereotypes. I investigate whether the impact of teacher IAT score on the gender gap is due to higher performance of boys, lower performance of girls, or a combination of the two effects. I estimate the following equation:

$$y_{icsy} = \beta_0 + \beta_1(Female_i \times stereotypes_c) + \beta_2 Female_i + \beta_3 stereotypes_c + \eta_{sy} + \\ + \mathbf{X}_i \rho_1 + (Female_i \times \mathbf{X}_i) \rho_2 + \mathbf{Z}_c \rho_3 + (Female_i \times \mathbf{Z}_c) \rho_4 + \varepsilon_{icsy} \quad (2)$$

where  $\eta_{sy}$  are school  $s$  by cohort  $y$  fixed effects and standard errors are robust and clustered at teacher level. All other variables are defined as in equation (1).

Institution level characteristics are captured in school by cohort fixed effects. The advantage with respect to specification (1) is that we can analyze the impact of teacher stereotypes separately on male students ( $\beta_3$ ) and on female students ( $\beta_1 + \beta_3$ ). The drawback is that I cannot control for unobservable characteristics at the teacher or class level: this specification exploits variation in the level of teacher stereotypes to which students of the same gender in the same school and cohort are exposed.

---

<sup>28</sup>I discuss the exogeneity of student assignment to teachers in Section 4.3.

<sup>29</sup>Glover et al. (2017), while analyzing the impact of manager implicit bias on minority workers, suggest that we may expect an attenuation bias of approximately a factor of 1.8 due to measurement error in the IAT score.

## 4.2 Gender Representativeness, IAT, and teachers' characteristics

Teacher gender stereotypes are driven by a *kernel-of-truth* (Bordalo et al., 2016): as in most countries, girls in Italy have lower standardized test scores in math and higher scores in reading compared to boys (Figure A.II). Despite the substantial overlap among distributions of ability, teachers may form inaccurate stereotypes by exaggerating the negative associations between math-female and reading-male. Figure A.II plots the representativeness of girls in each decile of the distribution of standardized test score in grade 8,  $\pi_{d,G}/\pi_{d,B}$ , where  $\pi_{d,G}$  is the probability of being in decile  $d$  for girls and  $\pi_{d,B}$  is the probability of being in decile  $d$  for boys (Bordalo et al., 2016). For example, girls are 1.6 times more likely than boys to be represented among the top 10% of the reading distribution, but 1.5 times less likely to be represented among students in the top 10% of the math ability distribution.<sup>30</sup> The presence of teachers' stereotypical associations between gender and field is consistent with the prediction of Bordalo et al. (2016): gender stereotypes amplify systematic differences between groups, ignoring the substantial overlap of the ability distributions of boys and girls.

The IAT captures implicit associations between *math-male* and *literature-female* (vs. *math-female* and *literature-male*): I cannot distinguish between the stereotype that *women are bad at math* and *men are bad at reading*. Figure I plots the entire distribution of implicit bias for math and literature teachers by gender: interestingly, individuals teaching a subject which is stereotypically associated with their own gender (i.e., men teaching math and women teaching literature) have stronger implicit male-math and female-literature associations. This result suggests that individuals possess implicit gender stereotypes in self-favourable form, likely because of the tendency to associate self with desirable traits – in this case, own gender with the subject they teach (Rudman et al., 2001).

The richness of the data collected allows me to explore the determinants related to the reaction

---

<sup>30</sup>In grade 10, the gender difference in math increases even more with two boys for every girl among the top 10% of math ability distribution in PISA 2015 data (Online Figure A.III). The gender stereotypical representativeness in math at the top and bottom of the ability distribution is substantially stronger in Italy compared to the US, where there are slightly less gender stereotypes (Nosek et al., 2009). In reading, there are no substantial differences among the two countries.

time to stimuli in the IAT score. Panel A of Table III shows that women teaching math have lower implicit stereotypes (column 1), but age, education of own mother, and whether teachers have children do not have a statistically significant correlation with IAT scores (columns 2-5). Gender stereotypical beliefs are rooted in cultural traits, transmitted from generation to generation (Guiso et al., 2006). I find that exposure to cultural norms is strongly associated with the IAT score. Column 1 of Panel B (Table III) shows that implicit stereotypes are correlated with the place of birth of teachers: around 40 percent of math teachers in this sample are born in the South where gender norms are stronger, as shown for instance by Campa et al. (2010).<sup>31</sup> I further investigate this aspect by providing evidence that female labor force participation in the province of origin of teachers is negatively correlated with IAT score (Panel B, column 2). As a proxy of cultural norms in the province of birth, I also use the answers to the World Value Survey question on the relative rights of men and women to paid jobs when jobs are scarce.<sup>32</sup> I find a positive correlation between less conservative gender norms measured by this question and IAT scores (Panel B, column 3). In the survey I administered, I asked the same question to teachers themselves and found a low, and indistinguishable from zero, correlation (Panel B, column 4). There may be social desirability bias in the self-reported measure when teachers are interviewed in the school. In column 5 of Panel B, I correlate implicit bias and explicit beliefs about innate differences in ability between men and women and find a weak, yet indistinguishable from zero, correlation in the expected direction. This result is not surprising in light of findings in social psychology that implicit stereotypes often differ from explicit and self-reported stereotypes (Lane et al., 2007; Nosek et al., 2002).

In Panel C, I correlate the IAT score with qualifications (type of degree and whether the degree was achieved *cum laude*) and other proxies of quality of teachers (tenure, being the professor in charge of math Olympiads in the school, and experience in teaching).<sup>33</sup> Point estimates are

---

<sup>31</sup>Italy is a country with low labor market participation of women, but substantial geographic variation across regions. In 2016, only 31 percent of women in the South of Italy were employed, while in the North around 58 percent were working, similar to the average of OECD.

<sup>32</sup>Thanks to the data used in Campa et al. (2010), I have access to the answers at province level of the following World Value Survey question: "When jobs are scarce, men have more right to a job than women".

<sup>33</sup>In each school, usually only one professor is in charge of math Olympiad and anecdotally the teacher is

small and indistinguishable from zero, with the exception of achieving a degree *cum laude*.<sup>34</sup> I also check whether the Gender-Science IAT score is correlated with the race IAT score. In the same regression as Table III, I find that the correlation is negative (-0.074 with standard error 0.123). Hence, math teachers more biased in one sphere are not more biased in the other sphere. The IAT score does not seem to capture a general “ability” in doing this type of test for math teachers.

Online Appendix Table A.IV shows all correlations presented in separate regressions in Table III together, for the sample of teachers whose data was matched with student outcomes (column 1) and for all teachers who completed the survey (column 2). Interestingly, the results are very similar. Finally, columns 3 and 4 of the online Appendix Table A.IV provide evidence of the correlation between characteristics of literature teachers and their IAT score. As shown in Figure I, female literature teachers are more likely than male literature teachers to associate math-male and literature-female. This is by far the most relevant factor in explaining the IAT score of literature teachers.

### 4.3 Exogeneity Assumption

Next, I present evidence on the absence of a systematic correlation between teacher gender stereotypes and student characteristics. If parents are able to guess which teachers have more stereotyping behaviour, they may try to (informally) affect class assignment of their daughters. Although this seems unlikely because implicit stereotypes are not an easily observable trait, it is also possible that parents try to select teachers according to characteristics correlated with IAT score, such as gender and place of birth. Furthermore, even if some parents manage to allocate their children to teachers with higher “quality”, it does not necessarily mean that they are less

---

highly motivated and passionate. Indeed, as shown in the Online Appendix Table A.III, teachers in charge of math Olympics induce higher improvements in test scores of their students. Similarly, teachers with tenure and more experience tend to have students with higher scores in standardized tests.

<sup>34</sup>There is a higher likelihood of obtaining a degree *cum laude* for teachers born in the South that may partially drive the correlation between IAT and degree with laude. This is a well-known fact in Italy: the share of students obtaining the degree *cum laude* is 28% in the North-West, 32% in the North-East, and 44% in the South (Source: MIUR data).

gender biased as shown in Table III.<sup>35</sup>

Table IV reports the correlation between student characteristics and stereotypes of math and literature teachers in Panel A and B, respectively. In column 1 of Panel A, I provide evidence that girls are not systematically assigned to math teachers with stronger or weaker gender stereotypes than boys, while in column 2 I show that daughters of highly educated mothers are not less likely to be assigned to teachers with more stereotypes than those from lower socioeconomic background – the difference is not statistically significant and the point estimate goes in the opposite direction. In columns 3 and 4 of Panel A, I analyze the correlation with father occupation and immigration background and I do not find a statistically significant correlation. The point estimates are also very small in terms of magnitude and the results are similar including all characteristics jointly (column 5). The p-value for the F-test of overall significance of these variables is 0.379, suggesting that we cannot reject the joint null-hypothesis at conventional levels. Finally, in the last column, I also include the standardized test score in math in grade 5 before entering middle school despite the sample size being reduced substantially due to data availability issues.<sup>36</sup> The assumption of “as good as random” assignment of students to math teachers with different IAT scores within a school seems to be supported in this context. Panel B reports the correlations between the same student characteristics and literature teacher stereotypes. In this case, some point estimates are statistically different from zero at conventional levels, even if they are small in magnitude and often in the opposite directions when all controls are jointly included in columns 5 and 6. The inclusion of these controls is potentially more relevant while analyzing the impact of literature teacher stereotypes. The results are identical when observations are collapsed at the teacher level, as shown in the Online Appendix Table A.V.

---

<sup>35</sup>In Italy, parents dislike being assigned to a teacher with a temporary contract who may have little experience and may change during the years of middle school. Teachers have a lot of experience (on average 22 years) and more than 90% have a full-time contract.

<sup>36</sup>Unfortunately, for reasons related to confidentiality, I only obtained the standardized test scores in grade 5 for those students who did not change school code between elementary and middle school. There are few students for whom I have this information and it is not a random sample: they are slightly more likely to be female and less likely to have highly educated mothers.

The second aspect regards the absence of a systematic grouping of students by socioeconomic background and initial ability. Within schools, classes are formed by the principal with the main objective of creating comparable groups in terms of gender, ability, and socioeconomic background across classes and, therefore, guarantee heterogeneity within each class in the same school and cohort. This objective is spelled out in the official documents on the school websites and also emerges from self-reported information from principals discussed in the Online Appendix B. I have information about the observable characteristics of students that are used to create classes (gender, education and occupation of parents, immigration status, and generation of immigration). Plausibly, unobservable student characteristics are also unknown to school principals at the moment of class formation, especially considering that students change all their teachers and their school building when moving from elementary to middle school. I check whether class assignments are statistically independent of student characteristics with a series of Pearson Chi-Square tests. First, I consider the assignment of individual level characteristics (gender, education and occupation of parents, immigration status, and generation of immigration). Then, I also check that, within each characteristic, class assignment is statistically independent from gender. I find that in less than 10% of the tests performed, the p-value is lower than or equal to 5%. There is no evidence of strong systematic grouping of students according to their socioeconomic background. In section 5.1.2, I provide evidence on the robustness of the main results. First, I do a permutation test where I randomly assign stereotypes to teachers. Second, I restrict the dataset to classes where assignment to peers is statistically independent for all student characteristics by gender.

### ***Timing of IAT collection***

Teachers' gender stereotypes were collected between October 2016 and March 2017, and they are matched with data of students who graduated from middle school between June 2012 and June 2017 (the detailed timeline is available in Figure II). An advantage of exploiting data of students who graduated before their teachers did the survey is that taking the IAT or knowledge

about this study could not have affected students' performance nor teachers' or parents' attention to the issue of gender stereotypes for the cohorts of students graduating before 2017. A potential concern is that IAT scores may be affected by exposure to the same cohorts of students. Indeed, the IAT is expected to be the combination of a trait stable over time capturing individual stereotypes and occasion-specific variation and noise that may be affected by conditions while taking the test and stimuli received by the subject in the period right before the test.<sup>37</sup>

Reverse causality seems unlikely for several reasons. I can provide supporting evidence against this issue by showing that the results are unchanged when I restrict the sample to the last cohort of students who graduated after their teachers took the IAT test (results are presented in Section 5.1.2). Furthermore, as shown in Section 4.3, teachers with more stereotypes are not systematically assigned to students with different characteristics, such as family background and standardized test scores in math (see Table IV and Online Appendix Table A.V). Teachers included in our analysis have been teaching, on average, for 20 years (with a median of 22 years) and, therefore, over time they were exposed to hundreds of female and male students.

## 5 The Impact of Teachers' Implicit Stereotypes

### 5.1 Performance in math

Table V shows the effect of teachers' implicit stereotypes on the gender gap in standardized test scores in grade 8 within the class, presenting the results of estimating equation (1). I document the impact of math and literature teachers in Panels A and B respectively. By the age of 14, girls lag 0.18 standard deviations behind in math compared to their male classmates (column 1 of Table V, Panel A).<sup>38</sup> Classes that are assigned to teachers with a one standard deviation

---

<sup>37</sup>The test-retest reliability of IAT is generally considered as satisfactory by social psychology, with a correlation of 0.56 that does not change with the length of time between testing (less than one month in most studies) (Nosek et al., 2007).

<sup>38</sup>This result is comparable to several other countries (Fryer and Levitt, 2010; Bharadwaj et al., 2016). In the Online Appendix Figure A.IV, I show the average gap in PISA test scores across countries. According to a meta-analysis performed on 100 studies in several countries, gender gaps in mathematics are around 0.29 standard deviations in high-school (Hyde et al., 1990), two years after the end of middle school. The average gender gap

higher IAT score during the three years of middle school have a 0.032 standard deviation larger gender gap in math performance in grade 8.<sup>39</sup> Column 3 includes student characteristics ( $\mathbf{X}_i$ ) and column 4 adds their interaction with gender of the children, without affecting the coefficient of interest.

Although the level of teacher stereotypes and all characteristics are absorbed by the class fixed effect, column 5 includes the interaction between student gender and teacher characteristics ( $\mathbf{Z}_c$ ). The magnitude and significance of the coefficient of interest (*Fem\*Teacher Stereotypes*) is not affected when all these interaction effects are absorbed. Observable characteristics of teachers, interacted with students' gender, are not driving the relation between gender gap and teacher stereotypes. I report the coefficients only for the main characteristics, but the effects are mainly small and insignificant at conventional levels for all variables, including age, parents' education, whether the teacher has daughters, whether he or she achieved the degree *cum laude*, the type of teaching contract, refresher courses, and appointment as teacher in charge of math Olympics. *Ceteris paribus*, female students assigned to female teachers have slightly (albeit insignificantly) higher math performance in grade 8 compared to their classmates<sup>40</sup>. The absence of a differential impact on boys and girls of teacher gender is consistent with the result of Bharadwaj et al. (2016). However, other studies find that having a teacher of the same gender helps improve performance, especially at the college level (Dee, 2005; Carrell et al., 2010). In the Online Appendix Table A.VI, I split the sample by teacher gender. The point estimate shows that the impact of teachers' implicit stereotypes on student performance is slightly larger in terms of magnitude for male, compared to female, teachers. However, what seems to matter the most is whether the teacher has gender stereotypes.

To give a clearer interpretation, columns 1-3 of the Online Appendix Table A.VII show the without controlling for class fixed effects is substantially invariant (0.18 standard deviations as shown in Table II). Most of the variation in math performance is within classes.

<sup>39</sup>These effects are related to exposure during a three-year period, with the exception of classes that changed teacher during middle school. The next section will focus on exposure for shorter time periods, exploiting data on standardized test scores in grade 6 when available.

<sup>40</sup>It should be noticed, however, that most teachers in Italian middle schools are females, in both math and literature. There is little variation in the gender of teachers and potentially substantial self-selection into the teaching profession, which differs by gender.

impact on the gender gap in the class of being assigned to a teacher with a positive (“boys-math” association) or a negative score (“girls-math” association) on the IAT test. The gender gap in the classroom is around 0.15 s.d. for students assigned to a teacher with an IAT score greater than zero, and it increases by one third (0.20 s.d) for classes assigned to a teacher with an IAT score lower than zero. In columns 4-6, I consider the thresholds defined by [Greenwald et al. \(2003\)](#), where “no stereotypes” is the interval of IAT raw score between -0.15 and +0.15, while “boys-math” and “girls-math” indicate a stronger association of the scientific field with male and female names respectively. Most of the difference in the gap is driven by being assigned to a teacher with a “boys-math” attitude or with “no systematic associations” (75% of teachers) compared to a teacher with a “girls-math” attitude (25% of teachers).

Are teachers with stronger stereotypes worse instructors or are they helping boys to learn math? Next, I investigate the effect of teacher bias from estimating equation (2) directly, comparing students of the same gender within the same school and cohort but assigned to different classes. [Figure III](#) shows that having a teacher with a strong “boys-math” attitude has a negative impact on female students, while a “girls-math” attitude has a positive impact on math improvements of girls. The linear approximation presented in [Table V](#) seems to adequately represent the data. There is no statistically significant impact on male students throughout the whole distribution of teachers’ IAT scores. Column 5 of [Table VI](#) mirrors [Figure III](#): it presents the results of the regression analysis and shows that girls are lagging behind when assigned to teachers with stronger implicit associations, while boys are not affected by teacher stereotypes. The results are robust to the inclusion of the controls as in [Table V](#). In this specification, the characteristics of teachers are not absorbed by class fixed effects and therefore controls at the teacher level are particularly relevant and column 5 is the preferred specification.

### **5.1.1 Heterogeneous effects: student ability and interaction time with teachers**

I examine which students are the most affected by teacher stereotypes, considering their background characteristics and the time of exposure to their teachers. [Table A.VIII](#) shows that the

effect of implicit stereotypes is stronger for female students who started middle school at the middle or lower-end of the initial ability distribution. Based on the estimates in column 2, a one standard deviation increase in teacher bias leads to -0.095 standard deviation lower performance among girls in the lowest tercile of test scores in grade 6, -0.068 (standard error 0.01) and +0.040 (standard error 0.22) for those girls in the middle and top terciles, compared to boys in the same initial tercile.<sup>41</sup> There are no significant heterogeneous effects according to other background characteristics, such as mothers' education or whether the student is an immigrant.<sup>42</sup>

Why do girls with lower level of ability initially suffer the most from the interaction with biased teachers? The empirical evidence presented is consistent with the stereotype threat model (Steele and Aronson, 1995): individuals with higher risk of conforming to the predicament that “*women are bad at math*” are those more deeply affected. Indeed, male students are not influenced by teacher stereotypes and, among females, those most strongly affected have lower initial math achievement and are at higher risk of confirming the negative expectations of their group. The Online Appendix E presents a conceptual framework that illustrates how teacher stereotypes can differentially affect effort and outcomes of students at the bottom and the top of the ability distribution.<sup>43</sup> One complementary explanation, consistent with the interaction theory (McConnell and Leibold, 2001), is that female students with higher initial math achievement may need less interaction with their math teacher in order to avoid lagging behind with their peers. This result is also consistent with evidence from Tiedemann (2002): teachers' perception of the math ability of their students is biased mainly toward average and low achieving female students who are perceived as less talented compared to their actual performance.

In order to investigate this further, I analyze the differential effect according to the quantity of interaction time between a teacher and their students. The last two columns of Table A.VIII

---

<sup>41</sup>Ideally, I should have created the terciles according to the test score in grade 5, before students were assigned to middle school teachers. This is available only for few students per class. I build the terciles using test scores in grade 6 for the cohort before 2013 because this test was not administered after that year.

<sup>42</sup>Although the point estimates are statistically indistinguishable from zero, the negative effect is bigger in magnitude for girls for whom I do not have official information on their parental background. This is more likely to happen for low performing students whose parents do not report information about own job and education to the school.

<sup>43</sup>This conceptual framework is an extension of the stereotype threat model presented by Dee (2014).

analyze whether there are heterogeneous effects in terms of years of exposure and hours per week.<sup>44</sup> Furthermore, I exploit the fact that around 25% of classes did not have the same teacher for all three years of middle school.<sup>45</sup> For both variables, I do not see a statistically significant pattern, but the point estimates suggest that longer exposure substantially increases the gender gap in the classroom. Girls after three years of exposure are lagging 0.037 standard deviations behind compared to their male classmates, while the effect is only 0.006 for those who changed teacher (column 6). Consistent with this result, Appendix Table A.IX shows the impact of one year or less of exposure to teachers' implicit stereotypes. I exploit two different samples. First, I use test score in grade 6, administered few months after assignment to middle school teachers (columns 1-3) and collected only up to 2012-13 and reported only for those teachers who took the test in 2017. Second, I exploit the fact that some classes were assigned to a new teacher at the beginning of grade 8 (columns 4-6). In both cases, the point estimates are indistinguishable from zero.

### 5.1.2 Robustness checks

First, even if all students are supposed to take the standardized test in grade 8, I do have missing information on the grade for 7% of students.<sup>46</sup> This mismatch may be due to some students not attending the test or schools misreporting the code to match data from the Ministry of Education and the code of the standardized test score with INVALSI. Online Table A.X shows that girls are around 2.5 percentage points less likely to have missing test scores than boys, but the effect is not statistically different for classes assigned to teachers with higher or lower level of stereotypes.

---

<sup>44</sup>Around 75% of students interact with the math teacher for six hours per week, while the rest interact for 9 hours per week.

<sup>45</sup>I observe the assignment of teachers to students since 2011. Hence, for the first two cohorts of students I do not know their teacher for at least one year. I assume they had the same teacher throughout middle school since their teachers have been working in the school for at least six consecutive years. The impact is similar when excluding these classes.

<sup>46</sup>Data are either present or missing for both test scores in math and literature, with the exception of 0.14% of cases.

Second, in the Online Appendix Table [A.XI](#), I show the effect of the main specification presented in Table [V](#) separately by cohorts of students who graduated before teachers took the IAT (school years 2012-2016) and for the cohort of students who graduated after teachers took the IAT (school year 2017). Reassuringly for the potential reverse causality concerns expressed in Section [4.3](#), results are statistically indistinguishable and the point estimate is larger for the last cohort of students.<sup>47</sup>

Third, Online Figure [A.V](#) plots the coefficient “Fem\*Stereotypes” from a permutation test that runs the main regression in equation (1) 1000 times randomly assigning the stereotypes to math teachers. In 5 out of 1000 permutations, I find a coefficient smaller than the one in Table [V](#). Finally, in the Online Table [A.XII](#), I restrict the sample to schools by cohorts where Pearson Chi-Square tests suggest statistical independence of all student characteristics (gender, education of the mother, occupation of the father, immigrant dummy, generation of immigration) and of all student characteristics by gender. Also, in this additional robustness check, results are not affected.

## 5.2 Performance in reading

Girls outperform boys in reading by 0.22 standard deviations (column 1 of Table [V](#), Panel B): the gender gap in female-typed areas is reversed compared to the one in male-typed areas discussed in the previous section, similarly to most OECD countries ([Fryer and Levitt, 2010](#)). Panel B of Table [V](#) focuses on the impact of literature teacher stereotypes on reading performance. Although the point estimate is negative, the gender stereotypes of literature teachers do not statistically significantly affect this gap. Panel B of the Online Appendix Table [A.VI](#) shows that the negative point estimate is mainly driven by male teachers, but, even for this subsample of teachers, the impact is not statistically significant at conventional levels.<sup>48</sup>

---

<sup>47</sup>Unfortunately, there are no cohorts of students exposed for the first time to teachers after they took the IAT. However, the fact that results are if anything stronger for the last cohort of students is reassuring for the potential reverse causality issue.

<sup>48</sup>The results are confirmed by the robustness checks reported in Panel B of the Online Figure [A.V](#) and the Online Table [A.XII](#).

The Online Appendix Table [A.XIII](#) investigates the impact of teacher stereotypes, considering the implicit IAT of both literature and math teachers and restricting the sample to those classes for which these scores are jointly available. The implicit stereotypes of literature teachers do not have a significant impact on math (columns 1-4) or on reading standardized test scores (columns 5-8). The inclusion of their IAT scores does not affect the negative and statistically significant effect of math teacher' stereotypes on math performance. Indeed, being assigned to a math teacher with stronger implicit stereotypes seems to have a negative, although indistinguishable from zero, effect on performance in reading, suggesting that female students do not simply substitute their effort in math for more effort devoted to studying literature.

There are several potential explanations for these results. First, it could be due to a measurement issue. Improvements in math may be easier to detect and measure on multiple choice tests. Standardized test scores in reading may be less elastic in capturing improvements during middle school, after basic literacy is completed, while standardized test scores in math may be closely related to specific learning during more recent school years. A significant impact on math standardized test scores accompanied by a no, or smaller, impact in reading is a common result in the literature ([Bettinger, 2012](#); [Levitt et al., 2016](#); [Carlana et al., 2019](#)). Furthermore, Gender-Science IAT scores do not allow to distinguish between the stereotype that *women are bad at math* and *men are bad at reading*. If the former association is more salient, this test may be better at detecting stereotypes in the scientific field and therefore it may have higher predictive power for math. Second, students may need less support and interaction with their math teachers compared to their literature teachers in order to perform well on the respective tests. Both math and literature teachers with stronger implicit associations may end up interacting less (in terms of quantity or quality) with students of the stigmatized group, which is consistent with findings on the role of interaction between managers and minority workers by [Glover et al. \(2017\)](#). However, only girls may be negatively affected because in math the support and explanation of teachers may be crucial for learning. Third, math skills are likely to be mainly taught in school while reading is more likely to be supplemented by parents or other caregivers

at home. Hence, teacher stereotypes may matter more in subjects almost exclusively taught by teachers versus other adults. Finally, consistent with Kugler et al. (2017) and Große and Riener (2010), girls may be more vulnerable to the gender stereotype that “*women are bad at math*” compared to boys exposed to the gender stereotype that “*men are bad at reading*”.<sup>49</sup> Girls tend to be less likely to believe that their good performance is due to their talent and more likely to believe that it is due to their effort. This is disproportionately true for math performance, as I show in the Online Appendix D.

### 5.3 Choice of High-School Track and Teachers’ Recommendation

#### *Background*

High-school track choice is the first crucial career decision in the Italian schooling system. There are three main types of high-school: academic, technical, and vocational.<sup>50</sup> Within the academic track, the scientific and classical sub-tracks are top-tier. As shown in Table II, there are substantial gender differences in the type of track selected: the preferred choices among girls are academic tracks related to psychology, languages, and art, while boys’ preferred choices are academic scientific and technical technological tracks.

Each family receives a formal letter from the school with the sub-track recommended by the teachers, mainly driven by math and literature teachers who interact the most with students at school.<sup>51</sup> Students and their families are free to choose their most-preferred track, with no constraints based on grades or teachers’ official track recommendation. Table II documents that girls are less likely, on average, to be recommended to the vocational track (7 percentage points) and scientific track (4 percentage points) than boys.<sup>52</sup> In the survey to teachers, I directly ask

---

<sup>49</sup>Furthermore, the type of task affects gender differences in the willingness to compete, with wider gaps for stereotypically male tasks (Niederle and Vesterlund, 2010; Große and Riener, 2010).

<sup>50</sup>Students in different tracks have, in most cases, little to no interaction during the school day.

<sup>51</sup>Teachers recommend students to a specific sub-track (e.g., scientific academic track). In 9% of the cases they give more than one sub-track recommendation and in 6% of the cases they broadly recommend an academic track.

<sup>52</sup>The effect is very similar when class fixed effects are included (column 1, Panel A and B of Table IX) and also when student level controls are added (column 3, Panel A and B of Table IX). If grade 8 standardized test scores in math are included in the regression with student level controls and class fixed effects, the gender gap in the

which factors they mainly consider when giving the track recommendation (see Appendix C.2 for the specific question). Both math and literature teachers consider motivation and interest as the most important factors, followed by grades given to students and involvement of parents in school activities. Given that girls tend to have higher academic motivation<sup>53</sup>, the evidence on fewer recommendations toward vocational school for girls is not surprising.

Appendix Table A.XIV shows that track choice and teachers' track recommendation are correlated for vocational high-school (columns 1-3) and scientific high-school (columns 4-6), with the latter correlation being substantially stronger. This table underlines some interesting gender differences: girls are 10 percentage points more likely to follow their teachers' recommendation to vocational track, which implies a 31% increase with respect to boys, while both genders are equally likely to follow a recommendation to scientific track.

In this section, I explore the impact of teacher stereotypes on track choice at the end of middle school using an ordered logit, and then I focus on the choice of the vocational track and scientific academic track using a linear probability model.<sup>54</sup>

## ***Results***

Panel A of Table VII reports fixed effects ordered logit estimates using the BUC estimator (Baetschmann et al., 2015), in which the dependent variable assumes value 1 for the vocational track, value 2 for intermediate tracks (technical and non top-tier academic), and value 3 for top-tier high-school (scientific and classical). These three categories are created grouping students according with the average test scores in grade 8 (before tracking), as can be clearly seen in Online Figure A.VI.<sup>55</sup> As shown in column 2, stereotypes of math teachers have a negative and probability of scientific track recommendation increases to -0.110 (s.e. 0.007), while the gender gap in vocational track recommendation decreases to -0.013 (s.e. 0.006).

<sup>53</sup>(for data on the lower academic motivation of boys with respect to girls, see Carlana et al., 2019)

<sup>54</sup>The track choice and teachers' track recommendation are not available for the cohort graduating in 2016-17.

<sup>55</sup>Students in the scientific and classical academic track have substantially higher average performance in grade 8 (before tracking) than those in other academic tracks, who perform similarly to students in the technical track. Students in the vocational track have substantially lower performance on average. The results are summarized in Online Figure A.VI.

statistically significant effect on the choice of a better high-school for girls, and the effect is unchanged by the inclusion of student- and teacher-level controls (columns 3 and 4, respectively). Although all students are supposed to take the test, those who go to high-school without taking the test are disproportionately represented among students enrolled in the vocational track.<sup>56</sup> When I restrict the sample to students who actually attended the standardized test in grade 8, the impact of teacher stereotypes on track choice is substantially smaller and insignificant at the conventional level (column 5). In column 6, I include the quadratic of the math standardized test score in grade 8, as a potential mediator given the results in section 5.1. The inclusion of these controls does not affect the point estimate.

The last two columns of Panel A (Table VII) provide evidence of the absence of a statistically significant impact on track choice of literature teachers. This result seems to support the idea discussed in Section 5.2 that girls may be more vulnerable to the gender stereotypes of their math teachers compared to boys to the gender stereotypes of their literature teachers. Panel B of Table VII documents a similar pattern in terms of magnitude for teachers' track recommendation, although estimates are less precise and indistinguishable from zero when controls are included.

To delve deeper into the choice of the field of study, I provide evidence on the effect of teacher stereotypes at the bottom (vocational track) and the top (scientific track) of the ability distribution using a linear probability model and following the same structure of Table VII.<sup>57</sup> Girls are slightly more likely to attend a vocational track (1.2 percentage points) than boys, while they are 8.9 percentage points less likely to attend a scientific track, as shown in column 1 of Table VIII, Panel A and B, respectively. This gap slightly increases when we include student level controls (column 3).<sup>58</sup> Consistently with the result in Table VII, math teacher stereotypes have a strong positive and statistically significant impact on the choice of vocational

---

<sup>56</sup>As discussed in Section 5.1.2 and Online Table A.X, teacher stereotypes do not have a statistically significant impact on the probability of taking the test in grade 8.

<sup>57</sup>The results are substantively invariant when I consider classical and scientific tracks jointly in the linear probability model.

<sup>58</sup>From column 4, student level controls are interacted with students' gender. It is not straightforward to interpret the coefficient of "Fem" in Table VIII.

track for girls, with respect to boys in the same class. One standard deviation higher teacher stereotypes increases the probability of attending a vocational track for girls (with respect to boys) by around 2 percentage points, which corresponds to an increase of 11.4% with respect to the mean probability of attending vocational training for girls. When I restrict the sample to students who took the standardized test score in grade 8, the point estimate decreases by around one third and is no longer statistically significant at the conventional level (column 5, Panel A). The effect is mainly driven by students who did not take the test score in grade 8. Column 6 shows that the inclusion of the squared polynomial of standardized test score absorbs most of the residual effect of math teacher stereotypes on the choice of vocational track. The last two columns of Table VIII show that literature teacher stereotypes have no significant effect on track choice.

Panel B of Table VIII reports the OLS estimates for the probability of attending a scientific track. The impact of both math and literature teacher stereotypes is statistically indistinguishable from zero, although the point estimate is negative and close to one percentage point for math teachers.<sup>59</sup>

Online Appendix Table A.XV shows the results estimating equation (2), with school by cohort, instead of class, fixed effects. Column 2 confirms the previous evidence of an impact on female students of math teacher stereotypes in terms of choice of vocational training. Part of the gender gap within class captured in the specification using class fixed effects is due to the lower probability of boys choosing the vocational track when assigned to teachers with stronger gender stereotypes.

Table IX provides evidence of a similar pattern compared to Table VIII in terms of magnitude of the effect of math and literature teacher stereotypes on track recommendation for vocational (Panel A) and scientific (Panel B) paths, although the effect is generally slightly smaller and less precisely estimated compared to the one on track choice. Track recommendation is a joint

---

<sup>59</sup>For the subsample of students for whom I have data on both math and literature teachers (column 7), the impact of teacher stereotypes on scientific track choice is negative and statistically significant. The impact on vocational track choice is also stronger.

decision of the math and literature teacher, so own bias may be attenuated. Online Appendix Table A.XVI includes school-by-cohort fixed effects and suggests that girls may be slightly less likely to be recommended to the scientific track, even if the result is not robust to the inclusion of all set of controls.

To sum up, math teacher stereotypes have a substantial impact on track choice mainly by inducing more girls to self-select into the vocational track. The effect is driven by students at the bottom of the ability distribution or with missing data on test scores. The impact on scientific track is negative, but generally indistinguishable from zero. The scientific track is chosen by females with high achievement test scores whose performance was not affected by teacher bias, as shown in the analysis of heterogeneous effects in Section 5.1. Girls at the top of the math ability distribution are likely to have other academic-oriented role models in addition to their math teacher and a lower vulnerability to gender stereotypes. The result on track choice mirrors an analogous difference in teachers' track recommendation, though it is smaller in magnitude and less precisely estimated. Literature teacher stereotypes have no significant effects on track choice of boys or girls. As discussed above, a potential explanation is that girls may be more vulnerable to the gender stereotypes of their math teachers compared to boys to the gender stereotypes of their literature teachers.

## **6 Discussion of a Potential Mechanism: Self-Confidence**

Self-confidence may play a crucial role in affecting performance, especially in “gender-incongruent areas”, such as math for girls or literature for boys (Coffman, 2014). According to findings in social psychology, the development of academic self-concept begins in childhood and is strongly influenced in the period after elementary school by stereotypes communicated by parents and teachers (Ertl et al., 2017). Students may believe that both their own signal of ability and the signal received by teachers carry relevant information. However, if the signal received from teachers is biased by gender stereotypes, female students, for example, may develop a lower

self-assessment of own ability in the scientific field and potentially invest less in their STEM education. This idea is consistent with the stereotype threat theory developed in social psychological literature (Steele and Aronson, 1995), according to which individuals at risk of confirming widely-known negative stereotypes reduce their confidence and underperform in fields in which their group is ability- stigmatized (Spencer et al., 1999). In the Online Appendix E, I present a conceptual framework that develops the intuition for the stereotype threat theory.<sup>60</sup>

Table X assesses the extent to which teacher stereotypes affect one's own assessment of ability, for a sample of around 800 students for whom I collected self-confidence measures.<sup>61</sup> I present results for self-confidence, defined as the own-assessment of ability when controlling for the standardized test score in grade 6, focusing on math in Panel A, on reading in Panel B, and on the average of all other subjects in Panel C.<sup>62</sup> As shown in column 1 (Panel A), girls are 9.2 percentage points less likely to consider themselves good at math (which corresponds to 11 percent lower probability than males). Female students are generally found to be more critical about their abilities in math than male students even if they have the same grades, as shown in PISA tests as well (OECD, 2015). However, girls are 4.1 percentage points more likely to consider themselves good in reading, but on average both are equally confident. This evidence supports the view that individuals process information about their own ability in a biased manner (Möbius et al., 2014). In classes assigned to math teachers with a one standard deviation higher IAT score, the gender gap in self-confidence increases by 4.8 percentage points.

---

<sup>60</sup>Despite the rich literature in social psychology about stereotype threat since the 1990s, economists have only recently began directly analyzing this phenomenon, finding partially contradictory evidence. One of the first steps taken in this direction has been Fryer et al. (2008), which finds no evidence of stereotype threat behavior in influencing women's performance in math, while Dee (2014) shows a substantial impact of activating negatively stereotyped identity (i.e., student-athlete) on test score performance.

<sup>61</sup>This measure of self-confidence is correlated with future educational choices of individuals. For instance, students with higher self-confidence in math are more likely to attend the scientific track, even controlling for standardized test score at the end of middle school.

<sup>62</sup>In Table X, I report the results using a dummy variable which assumes value 1 if the student reports to be good or mediocre and 0 if the student reports to be bad as the outcome. The point estimates are in the same direction but noisier and indistinguishable from zero when using an ordered logit. This result seems consistent with the previous evidence of a stronger impact on the bottom of the ability distribution. Finally, I have standardized test scores in grade 6 for the great majority of these students. Unfortunately, the test score in grade 5 is available only for around 20% of students. However, in Table A.IX, I provide evidence that the impact of teacher stereotypes on the standardized test score in grade 6 is close, and statistically indistinguishable, from zero.

Adding student level controls interacted with pupil gender does not substantially affect the point estimate of interest (columns 3 and 4, Panel A).

In Section 5.1, I provide evidence that the gender gap in math performance increases during middle school in classes assigned to a more biased teacher. Hence, in columns 5 and 6 of Table X, I also control for the mediating role of performance measured at the end of middle school in order to analyze whether gender gap in own assessment is merely due to different performance in grade 8. I find that gaps in self-confidence are only slightly reduced. Teacher stereotypes seems to have an additional impact on math self-confidence, on top of performance in standardized test score, that may have detrimental effects for investment choices in education and occupation.

In Panels B and C of Table X, I focus on the impact of teacher stereotypes on self-confidence in reading and all other subjects respectively. Girls have slightly higher self-confidence in literature, although the point estimate is indistinguishable from zero. One potential explanation is related to the framing of the question: students are asked to report whether they believe they are “good”, “mediocre”, or “bad” at each subject. They may want to avoid saying that they are “bad” at both the two crucial subjects (math and literature) and compensate the low self-confidence in math with higher self-assessment in reading. There is no impact on other subjects. The effects are substantively unchanged when controls at the individual level (column 3 and 4), at the teacher level (column 6), and for the standardized test score in grade 8 are included. Finally, in column 7 of each panel, I analyze the impact of both math and literature teacher stereotypes, while in column 8 I focus only on the impact of literature teacher stereotypes. Gender stereotypes of literature teachers slightly decrease the gender gap in self-confidence in reading, while they have no statistically significant effect on math and other subjects.

This result is important for at least two reasons. First, it shows that self-confidence is affected by social conditioning from teachers. Second, this is an important mechanism to understand the effect of teacher stereotypes on math performance and track choice of female students.

## 6.1 Additional Outcomes

### *Explicit Bias*

In Table [A.XVII](#) in the Online Appendix I consider the impact of teachers' reported beliefs on gender differences in innate math abilities on student outcomes. I find that classes assigned to a math teacher who believes there are gender differences in math ability have a substantially larger gender gap in math performance, in the same direction as the results reported by [Alan et al. \(2018\)](#). The impact of IAT score on student achievement is not significantly affected when I control for reported bias (column 4). This evidence seems to support the distinctiveness of implicit and explicit cognition ([Greenwald et al., 1998](#)) in the context of teacher gender stereotypes. Consistent with the results reported using the IAT, literature teachers' explicit beliefs about gender innate ability do not have a statistically significant effect on reading performance (columns 5 and 8).

### *Bias in Grading*

Previous literature has shown the importance of gender bias in grading (i.e., the gender difference in blindly graded standardized test scores and teacher-assigned grades) in affecting performance in math and university choice ([Lavy and Megalokonomou, 2017](#); [Lavy and Sand, 2018](#); [Terrier, 2016](#)). A natural question is whether implicit associations affect bias in grading of teachers. I have information only on grades given by teachers at the end of the semester. As shown in the Appendix Table [A.XVIII](#), girls get higher grades on average compared to boys in both math and literature when we control for the standardized test score in the same grade. Females assigned to teachers with more stereotypes get a slightly lower grade, but the effect is small and indistinguishable from zero. However, it should be considered that grades are a categorical variable from 2 to 10, where 6 is the pass grade. As shown in Figure [A.VII](#) in the Online Appendix, there is a high bunching at the pass grade, especially for math, and almost half of the students obtain the same grade in math. There is little variability in teacher-assigned

grades at the bottom of the distribution, where the effect of teacher stereotypes on standardized test scores is stronger.

Additional outcomes on retention rates are reported in the Online Appendix [F](#).

## 7 Conclusion

In most OECD countries, women outnumber men in tertiary education, but they are by far a minority in highly paid fields such as science, technology, engineering, and math, especially when excluding teaching careers. The prospects for change are not optimistic: according to 2015 PISA data, less than 5 percent of 15-years-old girls are planning to pursue a career in these fields on average in OECD countries compared to around 20 percent of boys. Culture and social conditioning have a strong impact on the development of skills and educational choices. This paper shows that the gender gap in math performance is substantially affected by teacher implicit stereotypes. Girls, especially those with lower initial skills, are lagging behind when assigned to teachers with stronger “math-males” and “literature-female” implicit associations. Boys, the group not ability-stigmatized in terms of math performance, are not affected by teacher stereotypes. The effects on reading are asymmetric and literature teacher stereotypes do not affect the gender gap in reading. Math teacher stereotypes influence high-school track choice, inducing more female students to attend an easier high-school. Furthermore, they foster low expectations about their own ability and lead to female under-confidence in male-typed domains. Indeed, girls are more likely to consider themselves bad at math at the end of middle school if they are assigned to a teacher with stronger stereotypes, even controlling for their ability measured by standardized test scores. These findings are consistent with a model whereby ability-stigmatized groups under-assess own ability and underperform fulfilling negative expectations about their achievements. Implicit associations can form an unintended and invisible barrier to equal opportunity.

These results raise the question of which kind of policies should be implemented in order

to alleviate the impact of gender stereotypes. The implicit stereotypes, measured by IAT score at this stage of development, should not be used to make high-stakes decisions, such as hiring or firing. IAT scores are educational tools to develop awareness of implicit preferences and stereotypes, and they should not have normative ground (Tetlock and Mitchell, 2009). However, one set of potential policies may be aimed at informing people about their own bias or training them in order to assure equal behavior toward all individuals, especially within the schooling context (Alesina et al., 2018). An alternative way to fight against the negative consequences of stereotypes is reducing vulnerability to these stereotypes by increasing self-confidence of girls in math or providing alternative role models – as done in the context of Indian elections, where exposure to female leaders weakens gender stereotypes in the home and public spheres (Beaman et al., 2009), or in schools, by offering alternative STEM role models or coding course for girls (Breda et al., 2018; Carlana and Fort, 2019). More research is needed to further investigate the impact of both types of policies.

## References

- Alan, S., S. Ertac, and I. Mumcu (2018). Gender stereotypes in the classroom and effects of achievement. *Review of Economics and Statistics*.
- Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018). Revealing Stereotypes: Evidence from immigrants in schools. *NBER Working Paper n25333*.
- Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. *Handbook of Labor Economics* 3, 3143–3259.
- Antecol, H., O. Eren, and S. Ozbeklik (2014). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics* 33(1), 63–89.
- Arkes, H. R. and P. E. Tetlock (2004). Attributions of implicit prejudice, or “would Jesse Jackson ‘fail’ the Implicit Association Test?”. *Psychological inquiry* 15(4), 257–278.
- Baetschmann, G., K. E. Staub, and R. Winkelmann (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178(3), 685–703.

- Barbieri, G., C. Rossetti, and P. Sestito (2011). The determinants of teacher mobility: Evidence using Italian teachers' transfer applications. *Economics of Education Review* 30(6), 1430–1444.
- Baron-Cohen, S. (2003). *The Essential Difference: Men, Women, and the Extreme Male Brain*. Allen Lane, London.
- Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: does exposure reduce bias? *Quarterly Journal of Economics* 124(4), 1497–1540.
- Bertrand, M., D. Chugh, and S. Mullainathan (2005). Implicit discrimination. *American Economic Review*, 94–98.
- Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. *Handbook of Economic Field Experiments*, Pages 309–393.
- Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics* 94(3), 686–698.
- Bettinger, E. P. and B. T. Long (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review* 95(2), 152–157.
- Bharadwaj, P., G. De Giorgi, D. Hansen, and C. Neilson (2016). The gender gap in mathematics: Evidence from low-and middle-income countries. *Economic Development and Cultural Change*.
- Blanton, H., J. Jaccard, J. Klick, B. Mellers, G. Mitchell, and P. E. Tetlock (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology* 94(3), 567.
- Bohren, J. A., A. Imas, and M. Rosenberg (2018). The dynamics of discrimination: Theory and evidence. *Working Paper*.
- Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *Quarterly Journal of Economics*.
- Bordalo, P., K. B. Coffman, N. Gennaioli, and A. Shleifer (2018). Beliefs about gender. *American Economic Review*.
- Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2018). Can female role models reduce the gender gap in science? Evidence from classroom interventions in French high schools. *Working Paper*.
- Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics* 129(3), 1409–1447.
- Campa, P., A. Casarico, and P. Profeta (2010). *Gender Culture and Gender Gap in Employment*, Volume 57. Oxford University Press.

- Card, D. and A. A. Payne (2017). High school choices and the gender gap in stem. *NBER Working Paper*.
- Carlana, M. and M. Fort (2019). Girls Code It Better. *Working Paper*.
- Carlana, M., E. La Ferrara, and P. Pinotti (2019). Goals and gaps: Educational careers of immigrant children. *Working Paper*.
- Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics* 125(3), 1101–1144.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593–2632.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review* 104(9), 2633–2679.
- Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *Quarterly Journal of Economics*.
- Cooper, H. M. and T. L. Good (1983). *Pygmalion grows up: Studies in the expectation communication process*. Longman Publishing Group.
- Corno, L., J. Burns, and E. La Ferrara (2018). Interaction, Stereotypes and Performance. Evidence from South Africa. *BREAD Working Paper n. 549*.
- Cvencek, D., A. N. Meltzoff, and A. G. Greenwald (2011). Math–gender stereotypes in elementary school children. *Child development* 82(3), 766–779.
- Dasgupta, N. and A. G. Greenwald (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology* 81(5), 800.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review* 95(2), 158–165.
- Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry* 52(1), 173–182.
- Else-Quest, N. M., J. S. Hyde, and M. C. Linn (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological bulletin* 136(1), 103.
- Ertl, B., S. Luttenberger, and M. Paechter (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology* 8.
- Ferrer-Esteban, G. (2011). Beyond the traditional territorial divide in the Italian education system. Effects of system management factors on performance in lower secondary school. *Working Paper*.

- Fiedler, K. and M. Bluemke (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology* 27(4), 307–316.
- Fryer, R. G. and S. D. Levitt (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 210–240.
- Fryer, R. G., S. D. Levitt, and J. A. List (2008). Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study. *American Economic Review* 98(2), 370–375.
- Giustinelli, P. (2016). Group decision making with uncertain outcomes: Unpacking child–parent choice of the high school track. *International Economic Review* 57(2), 573–602.
- Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores. *Quarterly Journal of Economics*.
- Goldin, C., L. F. Katz, and I. Kuziemko (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives* 20(4), 133–156.
- Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology* 74(6), 1464.
- Greenwald, A. G., B. A. Nosek, and M. R. Banaji (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology* 85(2), 197.
- Greenwald, A. G., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology* 97(1), 17.
- Große, N. D. and G. Riener (2010). Explaining gender differences in competitiveness: gender-task stereotypes. *Jena Economic Research Papers*.
- Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Culture, gender, and math. *Science* 320(5880), 1164–1165.
- Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives* 20(2), 23–48.
- Guryan, J. and K. K. Charles (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *Economic Journal* 123(572), F417–F432.
- Hyde, J. S., E. Fennema, and S. J. Lamon (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin* 107(2), 139.
- Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

- Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1), 5–86.
- Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology* 141(2), 165–173.
- Kiefer, A. K. and D. Sekaquaptewa (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology* 43(5), 825–832.
- Kugler, A. D. et al. (2017). Choice of majors: Are women really different from men? *CEPR Discussion Papers*.
- Lane, K. A., M. R. Banaji, B. A. Nosek, and A. G. Greenwald (2007). *Understanding and using the Implicit Association Test: IV*. New York: Guilford.
- Lavy, V. and R. Megalokonomou (2017). Persistency in teachers' grading biases and effect on longer term outcomes: University admission exams and choice of field of study. *Working Paper*.
- Lavy, V. and E. Sand (2018). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases. *Journal of Public Economics*.
- Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy* 8(4), 183–219.
- McConnell, A. R. and J. M. Leibold (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology* 37(5), 435–442.
- Meghir, C. and M. Palme (2005). Educational reform, ability, and family background. *American Economic Review* 95(1), 414–424.
- Möbius, M., P. Niehaus, M. Niederle, and T. Rosenblat (2014). Managing self-confidence. *Working Paper*.
- Murnane, R. J., J. B. Willett, and F. Levy (1995). The growing importance of cognitive skills in wage determination. Technical report, National Bureau of Economic Research.
- Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives* 24(2), 129–144.
- Nollenberger, N., N. Rodríguez-Planas, and A. Sevilla (2016). The math gender gap: The role of culture. *American Economic Review* 106(5), 257–261.
- Nosek, B. A., M. R. Banaji, and A. G. Greenwald (2002). Math= male, me= female, therefore math $\neq$  me. *Journal of Personality and Social Psychology* 83(1), 44.

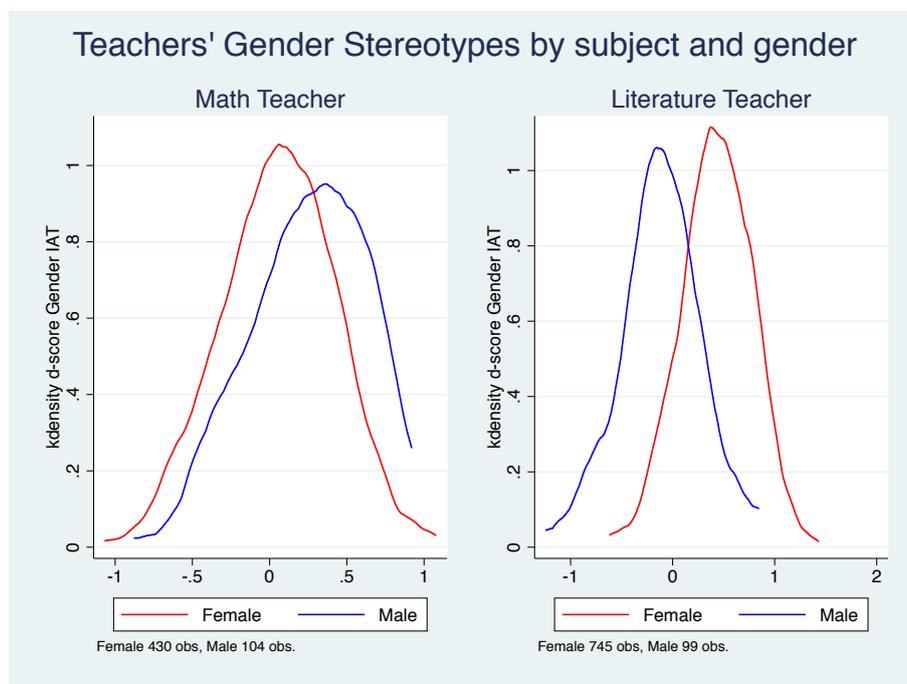
- Nosek, B. A., F. L. Smyth, J. J. Hansen, T. Devos, N. M. Lindner, K. A. Ranganath, C. T. Smith, K. R. Olson, D. Chugh, A. G. Greenwald, et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology* 18(1), 36–88.
- Nosek, B. A., F. L. Smyth, N. Sriram, N. M. Lindner, T. Devos, A. Ayala, Y. Bar-Anan, R. Bergh, H. Cai, K. Gonsalkorale, et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106(26), 10593–10597.
- OECD (2014). Are boys and girls equally prepared for life?
- OECD (2015). The ABC of gender equality in education: Aptitude, behaviour, confidence.
- Olson, M. A. and R. H. Fazio (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *Journal of Personality and Social Psychology* 86(5), 653.
- Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology* 105(2), 171.
- Papageorge, N. W., S. Gershenson, and K. Kang (2018). Teacher expectations matter. *NBER Working Paper*.
- Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences* 111(12), 4403–4408.
- Reuben, E., M. Wiswall, and B. Zafar (2015). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *Economic Journal*.
- Riegle-Crumb, C. and M. Humphries (2012). Exploring bias in math teachers’ perceptions of students’ ability by gender and race/ethnicity. *Gender & Society* 26(2), 290–322.
- Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics* 17(3), 523–534.
- Rosenthal, R. and L. Jacobson (1968). Pygmalion in the classroom. *Urban Review* 3(1), 16–20.
- Rudman, L. A., A. G. Greenwald, and D. E. McGhee (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin* 27(9), 1164–1178.
- Spencer, S. J., C. M. Steele, and D. M. Quinn (1999). Stereotype threat and women’s math performance. *Journal of experimental social psychology* 35(1), 4–28.
- Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69(5), 797.
- Terrier, C. (2016). Boys Lag Behind: How Teachers’ Gender Bias Affect Students’ Achievement. *SEII Discussion Paper*.

Tetlock, P. E. and G. Mitchell (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in organizational behavior* 29, 3–38.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics* 50(1), 49–62.

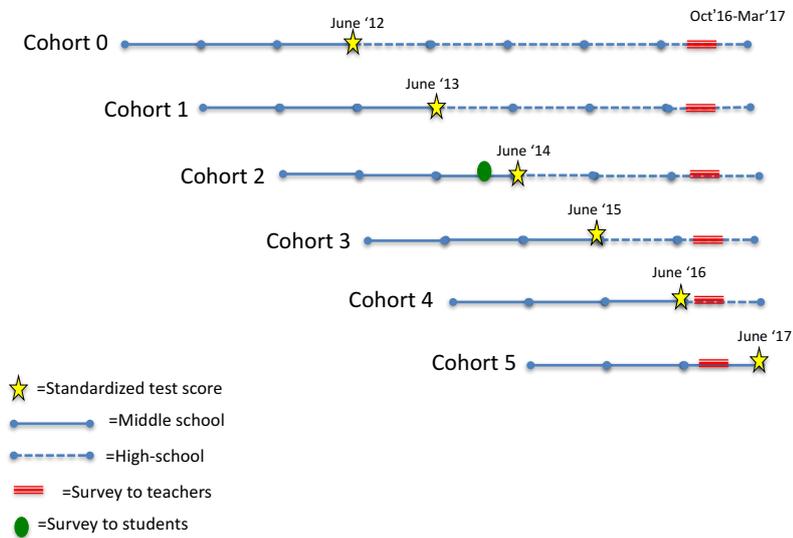
## Figures and Tables

Figure I. Teachers' Implicit Gender Bias (IAT measure) by gender and subject they teach



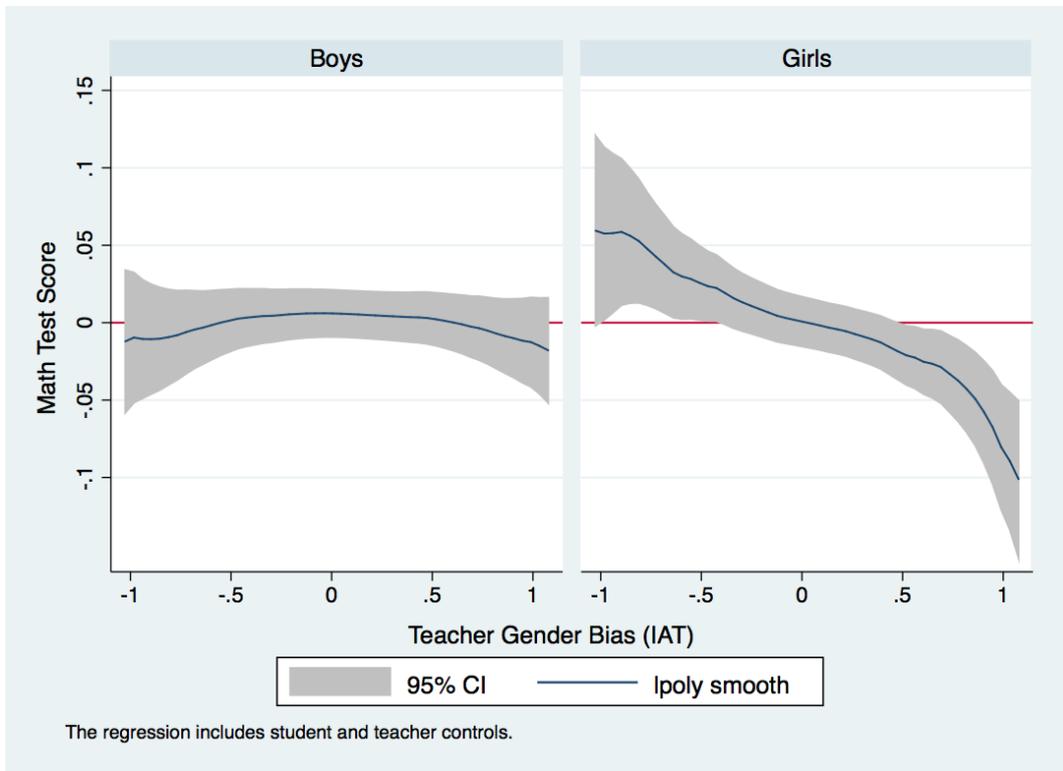
*Notes:* This graph shows the distribution of Gender-Science IAT scores for math and literature teachers, separated by gender. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes.

Figure II. Timeline of main data available for students and teachers



*Notes:* This graph shows the timeline of data collected for the three cohorts of students. They graduated from middle school between 2012 and 2017. Teachers were surveyed between October 2016 and March 2017. Standardized tests are administered at the end of grade 8.

Figure III. Effect of teacher bias on student math performance by gender



*Notes:* This graph shows the effect of teacher stereotypes on student achievement by gender. The variable on the y axis is the residualized standardized test score in grade 8, after controlling for school by cohort fixed effects, and student- and teacher-level controls. The variable on the x axis is the raw IAT score. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females.

Table I. Summary Statistics from Teachers' Questionnaire

	Count	Mean	SD	Min	Max
<b>Panel A: Math teachers</b>					
<b>Family and education</b>					
Female	454	0.81	0.39	0.00	1.00
Born in the North	440	0.61	0.49	0.00	1.00
Age	439	48.81	9.66	25.00	66.00
Children	454	0.68	0.47	0.00	1.00
Number of children	298	1.88	0.85	0.00	5.00
Number of daughters	298	0.87	0.77	0.00	4.00
Low edu Mother	417	0.52	0.50	0.00	1.00
Middle edu Mother	417	0.33	0.47	0.00	1.00
High edu Mother	417	0.15	0.36	0.00	1.00
Advanced STEM	442	0.23	0.42	0.00	1.00
Degree Laude	388	0.21	0.41	0.00	1.00
<b>Job characteristics</b>					
Full time contract	434	0.79	0.41	0.00	1.00
Years of experience	433	18.83	12.05	0.00	48.00
Math Olympiad	442	0.16	0.37	0.00	1.00
refresher courses	442	0.91	0.28	0.00	1.00
<b>Implicit and explicit stereotypes</b>					
IAT Gender	454	0.09	0.37	-1.03	1.08
WVS Gender Equality	438	0.16	0.36	0.00	1.00
No Gender Dif Innate Ability	422	0.85	0.36	0.00	1.00
<b>Panel B: Literature teachers</b>					
<b>Family and education</b>					
Female	615	0.90	0.30	0.00	1.00
Born in the North	591	0.72	0.45	0.00	1.00
Age	589	49.55	8.33	25.00	66.00
Children	615	0.71	0.46	0.00	1.00
Number of children	425	1.78	0.82	0.00	5.00
Number of daughters	425	0.84	0.76	0.00	4.00
Low edu Mother	547	0.49	0.50	0.00	1.00
Middle edu Mother	547	0.37	0.48	0.00	1.00
High edu Mother	547	0.14	0.34	0.00	1.00
<b>Job characteristics</b>					
Full time contract	594	0.94	0.23	0.00	1.00
Years of experience	591	21.67	10.25	0.00	43.00
refresher courses	606	0.93	0.26	0.00	1.00
<b>Implicit and explicit stereotypes</b>					
IAT Gender	615	0.38	0.39	-1.08	1.43
No Gender Dif Innate Ability	579	0.89	0.31	0.00	1.00
WVS Gender Equality	588	0.11	0.31	0.00	1.00

*Notes:* First-hand data from teachers' questionnaire. I restrict the sample to teachers matched to students and therefore used in the main analysis of this paper. The balance table with the difference between teachers matched and not matched with student data is presented in Table A.II.

Table II. Summary Statistics of students by gender

	Males	Females	Diff.	Norm. Diff.
<b>Baseline characteristics</b>				
Std Math grade 6	0.192 (1.014)	0.005 (0.950)	-0.188 (0.021)***	-0.135
Std reading grade 6	0.040 (0.973)	0.179 (0.892)	0.140 (0.020)***	0.106
Immigrant	0.208 (0.406)	0.201 (0.401)	-0.007 (0.005)	-0.013
Second Gen. Immigrant	0.089 (0.285)	0.090 (0.287)	0.001 (0.004)	0.002
HighEduMother	0.423 (0.494)	0.422 (0.494)	-0.001 (0.006)	-0.002
Missing Edu Mother	0.235 (0.424)	0.230 (0.421)	-0.005 (0.005)	-0.008
High Occupation Father	0.162 (0.368)	0.165 (0.371)	0.003 (0.004)	0.005
Medium Occupation Father	0.298 (0.457)	0.293 (0.455)	-0.005 (0.005)	-0.007
Missing Occupation Father	0.256 (0.437)	0.255 (0.436)	-0.001 (0.005)	-0.001
<b>Outcomes</b>				
Std Math grade 8	0.120 (1.005)	-0.062 (0.974)	-0.182 (0.012)***	-0.130
Std reading grade 8	-0.084 (0.998)	0.129 (0.974)	0.213 (0.013)***	0.153
High-school Track: Scientific	0.301 (0.459)	0.201 (0.401)	-0.100 (0.008)***	-0.164
High-school Track: Classic	0.036 (0.186)	0.074 (0.262)	0.038 (0.004)***	0.119
High-school Track: Other Academic	0.090 (0.286)	0.333 (0.471)	0.243 (0.007)***	0.441
High-school Track: Technical Technological	0.313 (0.464)	0.069 (0.253)	-0.244 (0.008)***	-0.461
High-school Track: Technical Economic	0.120 (0.325)	0.166 (0.372)	0.046 (0.006)***	0.092
High-school Track: Vocational	0.140 (0.347)	0.157 (0.364)	0.017 (0.006)***	0.034
Track recommendation: Scientific	0.216 (0.412)	0.177 (0.382)	-0.039 (0.007)***	-0.070
Track recommendation: Vocational	0.389 (0.488)	0.318 (0.466)	-0.071 (0.009)***	-0.106
Average own ability	0.654 (0.176)	0.641 (0.160)	-0.013 (0.010)	-0.054
Own ability: math	0.834 (0.372)	0.756 (0.430)	-0.078 (0.022)***	-0.138
Own ability: reading	0.919 (0.273)	0.964 (0.187)	0.045 (0.016)***	0.135
Observations	15,373	14,986	30,359	

*Notes:* This table reports the summary statistics and the difference between the two genders in outcomes and baseline characteristics. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively. The normalized difference shown in column 4 is the formula recommended by Imbens and Wooldridge (2009). More details are reported in footnote 11 of the paper.

Table III. Correlation between teachers' characteristics and Gender IAT Score

<b>Panel A: Independent variables (background teachers' characteristics)</b>					
	Female (1)	Age (2)	HighMotherEdu (3)	Children (4)	Daughters (5)
<b>Dep. Var.:</b>					
<b>Raw IAT</b>	-0.174*** (0.051)	-0.015 (0.020)	0.011 (0.035)	-0.072 (0.105)	0.035 (0.047)
Obs.	454	454	454	454	454
$R^2$	0.043	0.014	0.011	0.011	0.012
<b>Panel B: Independent variables (cultural traits and beliefs)</b>					
	BornNorth (1)	WomenLFP (2)	WVSCityBorn (3)	WVSIndiv (4)	InnateAbility (5)
<b>Dep. Var.:</b>					
<b>Raw IAT</b>	-0.081** (0.035)	-0.295** (0.146)	0.307*** (0.110)	-0.003 (0.047)	-0.028 (0.046)
Obs.	454	433	389	454	454
$R^2$	0.022	0.021	0.022	0.011	0.011
<b>Panel C: Independent variables (education and teacher experience)</b>					
	STEM (1)	Laude (2)	FullContract (3)	Olympiad (4)	HighExp (5)
<b>Dep. Var.:</b>					
<b>Raw IAT</b>	-0.060 (0.045)	-0.082** (0.039)	-0.075 (0.053)	0.067 (0.069)	-0.016 (0.063)
Obs.	454	454	454	454	454
$R^2$	0.016	0.019	0.019	0.200	0.017

*Notes:* This table reports OLS estimates of the correlation between math teachers' IAT score and own characteristics. The unit of observation is teacher  $t$  in school  $s$ . Standard errors (in parentheses) are robust and clustered at the school level. The number of clusters is 90. School fixed effects are included in all regressions. The significance and magnitude of coefficients are not affected by the inclusion of FE. The variable "Female" indicates the gender of the teacher, "Born in the North" assumes value 1 if the teacher was born in the North of Italy, "HighMotherEdu" is a dummy which assumes value 1 if the mother of the teacher has at least a diploma, "Children" and "Daughters" are dummies which assume a value of 1 if the teacher has children/daughters. The variable "STEM" assumes value 1 if the teacher has a degree in math, engineering, or physics; "Laude" is a dummy which assumes value 1 if the degree was achieved cum laude, "Full Contract" assumes value 1 if the teacher has tenure, "Olympiad" is 1 for teachers in charge of math Olympiad in the school; "HighExp" is a dummy variable which assumes value 1 if the teacher has more than 15 years of experience; "WomenLFP" is the labor force participation of women in the province of birth; "WVSCityBorn" is the WVS answer to the relative rights of men and women to paid jobs when jobs are scarce; "WVSIndiv" is the answer to the same question at the individual level, "InnateAbility" regards the teacher belief about innate differences in math abilities between men and women (1 means no differences in innate ability, 0 otherwise). I include the order of IATs for math teachers and missing categories if the information is not available. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table IV. Exogeneity of assignment of students to math teachers with different bias

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A- Dependent Variable: implicit gender stereotypes of math teacher (standardized)</b>						
Female	-0.007 (0.007)	-0.015 (0.011)	-0.015 (0.013)	-0.009 (0.008)	-0.023 (0.016)	0.003 (0.026)
HighEduMother		-0.023 (0.015)			-0.022 (0.014)	-0.037* (0.022)
Fem*HighEduMother		0.014 (0.018)			0.014 (0.019)	0.009 (0.036)
Medium Occupation Father			-0.015 (0.015)		-0.009 (0.014)	0.043 (0.031)
Fem*Medium Occupation Father			0.012 (0.020)		0.011 (0.020)	-0.041 (0.042)
High Occupation Father			-0.028 (0.021)		-0.018 (0.019)	0.023 (0.039)
Fem*High Occupation Father			0.002 (0.021)		-0.001 (0.024)	0.026 (0.059)
Immigrant				0.008 (0.018)	0.006 (0.018)	0.033 (0.030)
Fem*Immigrant				0.009 (0.019)	0.011 (0.020)	-0.004 (0.038)
Std Math 5						-0.013 (0.013)
Fem*Std Math 5						-0.019 (0.020)
Obs.	30359	30359	30359	30359	30359	6847
R <sup>2</sup>	0.338	0.338	0.338	0.338	0.338	0.348

Table IV. Exogeneity of assignment of students to literature teachers with different bias (*cont.*)

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel B- Dependent Variable: implicit gender stereotypes of literature teacher (standardized)</b>						
Female	0.014*	0.018	-0.008	0.008	-0.014	0.026
	(0.008)	(0.011)	(0.015)	(0.009)	(0.016)	(0.033)
HighEduMother		0.004			0.009	-0.021
		(0.014)			(0.014)	(0.024)
Fem*HighEduMother		-0.015			-0.026	-0.073**
		(0.017)			(0.018)	(0.036)
Medium Occupation Father			-0.031*		-0.039**	-0.013
			(0.017)		(0.018)	(0.034)
Fem*Medium Occupation Father			0.030		0.047**	0.038
			(0.022)		(0.023)	(0.046)
High Occupation Father			-0.007		-0.018	0.011
			(0.023)		(0.023)	(0.037)
Fem*High Occupation Father			0.036		0.058**	0.053
			(0.026)		(0.028)	(0.058)
Fem*Immigrant				0.031	0.041**	-0.007
				(0.020)	(0.020)	(0.042)
Immigrant				-0.017	-0.019	0.018
				(0.020)	(0.020)	(0.030)
Std Reading 5						0.007
						(0.012)
Fem*Std Reading 5						0.002
						(0.016)
Obs.	29486	29486	29486	29486	29486	6873
R <sup>2</sup>	0.344	0.345	0.345	0.345	0.345	0.417

*Notes:* This table reports OLS estimates of the correlation between teacher stereotypes measured by IAT score and student characteristics. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable “Fem” indicates the gender of the student, “HighEduMother” assumes value 1 if the mother has at least a 5-years high-school diploma, “Medium Occupation Father” assumes value 1 if the father is a teacher or office worker, while “High Occupation Father” is 1 if the father is a manager, university professor, or executive. “Immigrant” assumes value 1 if the student is not an Italian citizen, while “Std Math/Reading 5” is the standardized test score in grade 5 in mathematics/reading. All regressions include controls for the order of IAT in the questionnaire administered. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. The F-test for joint significance of all characteristics is 0.379 for Panel A (math teachers, column 5) and 0.054 for Panel B (literature teachers, column 5). \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table V. Estimation of the effect of teachers' gender stereotypes on standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)
<b>Panel A- Dependent Variable: Math standardized test score in grade 8</b>					
Female	-0.184*** (0.012)	-0.198*** (0.013)	-0.198*** (0.012)	-0.174*** (0.025)	-0.032 (0.082)
Fem*Math Teacher Stereotypes		-0.032** (0.013)	-0.031*** (0.012)	-0.032*** (0.012)	-0.031*** (0.011)
Fem* Teacher Fem					0.020 (0.030)
Fem* Teacher. Born North					0.010 (0.024)
Fem*Advanced STEM Teacher					-0.006 (0.024)
Gender Gap	-0.185	-0.185	-0.186	-0.186	-0.1856
Obs.	30359	30359	30359	30359	30359
R <sup>2</sup>	0.206	0.206	0.275	0.275	0.276
<b>Panel B- Dependent Variable: Reading standardized test score in grade 8</b>					
Female	0.221*** (0.012)	0.225*** (0.013)	0.221*** (0.012)	0.226*** (0.024)	0.307*** (0.093)
Fem*Lit Teacher Stereotypes		-0.012 (0.013)	-0.006 (0.012)	-0.007 (0.012)	-0.001 (0.013)
Fem* Teacher Fem					-0.044 (0.039)
Fem* Teacher. Born North					0.011 (0.027)
Gender Gap	0.221	0.221	0.219	0.219	0.219
Obs.	29486	29486	29486	29486	29486
R <sup>2</sup>	0.181	0.181	0.291	0.291	0.291
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math or reading standardized test score in grade 8 in Panel A and B, respectively, and the dependent variables refer to math teachers in Panel A and to literature teachers in Panel B. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level (454 in Panel A and 615 in Panel B). The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table VI. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - school FE regression

<b>Dependent Variable: Math standardized test score in grade 8</b>					
	(1)	(2)	(3)	(4)	(5)
Female	-0.180*** (0.012)	-0.192*** (0.013)	-0.196*** (0.012)	-0.166*** (0.024)	-0.047 (0.079)
Fem*Teacher Stereotypes		-0.029** (0.012)	-0.029** (0.011)	-0.031*** (0.011)	-0.030*** (0.011)
Teacher Stereotypes		-0.018 (0.013)	-0.015 (0.012)	-0.014 (0.012)	-0.013 (0.012)
Teacher Fem					0.066** (0.033)
Fem* Teacher Fem					0.018 (0.030)
Teacher Born North					-0.006 (0.025)
Fem* Teacher Born North					0.012 (0.023)
Advanced STEM					-0.048* (0.027)
Fem*Advanced STEM Teacher					-0.006 (0.024)
Gender Gap	-0.180	-0.180	-0.184	-0.184	-0.184
School, year FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls *Fem	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	30359	30359	30359	30359	30359
$R^2$	0.128	0.129	0.208	0.208	0.214

*Notes:* This table reports OLS estimates of equation 2, where the dependent variable is math standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The number of clusters is 454. The number of fixed effects (school by cohort) is 459. The variable “Fem” indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, education of the teacher’s mother and the interaction with student gender of all these characteristics. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table VII. Estimation of the effect of teachers' gender stereotypes on high-school track choice - ordered logit with class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: High-School Track Choice</b>								
Fem	-0.270*** (0.028)	-0.300*** (0.030)	-0.306*** (0.030)	-0.458** (0.219)	-0.483** (0.231)	-0.496** (0.237)	-0.370*** (0.048)	-0.292*** (0.032)
Fem*Math Teacher Stereotypes		-0.074** (0.031)	-0.070** (0.030)	-0.059* (0.030)	-0.035 (0.032)	-0.035 (0.034)	-0.094** (0.042)	
Fem*Lit Teacher Stereotypes							0.028 (0.045)	0.011 (0.031)
Individuals	21027	21023	20750	20725	17899	17016	11192	20254
Obs.	47351	47351	47351	47351	41772	41772	25395	45167
Pseudo-log-likelihood	-21026	-21022	-20749	-20724	-17896	-17012	-11191	-20029
<b>Panel B- Dependent Variable: High-School Track Recommendation</b>								
Fem	0.179*** (0.032)	0.151*** (0.035)	0.148*** (0.036)	0.137 (0.269)	0.201 (0.281)	0.406 (0.301)	0.063 (0.054)	-0.190*** (0.036)
Fem*Math Teacher Stereotypes		-0.068** (0.031)	-0.046 (0.035)	-0.059 (0.037)	-0.059 (0.037)	-0.046 (0.044)	-0.077 (0.048)	
Fem*Lit Teacher Stereotypes							0.041 (0.051)	0.006 (0.036)
Individuals	18684	18684	18684	18684	16904	16904	10575	17979
Obs.	39349	39349	39349	39349	34215	34215	21789	38048
Pseudo-log-likelihood	-17042	-17039	-15241	-15222	-13547	-10375	-9711	-16623
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	No	Yes	No	No

Notes: This table reports fixed effects ordered logit estimates (BUC estimators following Baetschmann et al. (2015)), where the dependent variable assumes value 1 for vocational track, value 2 for intermediate tracks (technical and no top-tier academic), and value 3 for top-tier high-school (scientific and classical). The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students with a standardized test score in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. "Sq. Math Test 8" is the second order polynomial of test score in grade 8. "Obs." denotes the number of student-choices in the estimation sample; "Individuals" denotes the number of unique people in the estimation sample. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table VIII. Estimation of the effect of teachers' gender stereotypes on high-school track choice - class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: Vocational High-School Track Choice</b>								
Fem	0.012** (0.006)	0.020*** (0.007)	0.019*** (0.006)	0.054 (0.047)	0.038 (0.047)	0.028 (0.045)	0.026*** (0.010)	0.011* (0.006)
Fem*Math Teacher Stereotypes		0.019*** (0.007)	0.016** (0.007)	0.013** (0.007)	0.008 (0.007)	0.004 (0.006)	0.023** (0.009)	
Fem*Lit Teacher Stereotypes							0.009 (0.009)	-0.001 (0.006)
Constant	0.163*** (0.003)	0.163*** (0.003)	0.190*** (0.008)	0.162*** (0.010)	0.180*** (0.010)	0.159*** (0.010)	0.153*** (0.004)	0.160*** (0.003)
Mean Y Fem	0.175	0.175	0.202	0.202	0.202	0.202	0.172	0.171
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R <sup>2</sup>	0.116	0.117	0.161	0.164	0.154	0.220	0.116	0.116
<b>Panel B- Dependent Variable: Scientific High-School Track Choice</b>								
Fem	-0.089*** (0.007)	-0.093*** (0.008)	-0.091*** (0.008)	-0.003 (0.051)	-0.012 (0.054)	-0.006 (0.051)	-0.107*** (0.012)	-0.092*** (0.008)
Fem*Math Teacher Stereotypes		-0.009 (0.007)	-0.006 (0.007)	-0.008 (0.007)	-0.008 (0.008)	-0.006 (0.007)	-0.024** (0.010)	
Fem*Lit Teacher Stereotypes							0.002 (0.011)	-0.006 (0.007)
Constant	0.283*** (0.003)	0.283*** (0.003)	0.196*** (0.008)	0.164*** (0.010)	0.160*** (0.011)	0.122*** (0.010)	0.281*** (0.005)	0.284*** (0.004)
Mean Y Fem	0.194	0.194	0.107	0.107	0.108	0.108	0.185	0.191
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R <sup>2</sup>	0.117	0.117	0.152	0.159	0.161	0.273	0.111	0.116
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	No	Yes	No	No

Notes: This table reports OLS estimates of equation 1, where the dependent variable is the high-school track choice. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. "Sq. Math Test 8" is the second order polynomial of test score in grade 8. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table IX. Estimation of the effect of teachers' gender stereotypes on high-school track recommendation of teachers - class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: Vocational High-School Track Recommendation</b>								
Fem	-0.074*** (0.008)	-0.067*** (0.009)	-0.062*** (0.008)	-0.074 (0.062)	-0.087 (0.064)	-0.110* (0.057)	-0.053*** (0.013)	-0.082*** (0.009)
Fem*Math Teacher Stereotypes		0.018** (0.009)	0.011 (0.008)	0.012 (0.008)	0.011 (0.009)	0.004 (0.008)	0.018 (0.012)	
Fem*Lit Teacher Stereotypes							-0.000 (0.013)	0.006 (0.009)
Constant	0.433*** (0.004)	0.433*** (0.004)	0.517*** (0.009)	0.516*** (0.012)	0.528*** (0.013)	0.470*** (0.011)	0.423*** (0.005)	0.429*** (0.004)
Mean Y Fem	0.359	0.359	0.450	0.450	0.450	0.450	0.362	0.349
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R <sup>2</sup>	0.143	0.143	0.271	0.272	0.244	0.427	0.132	0.138
<b>Panel B- Dependent Variable: Scientific High-School Track Recommendation</b>								
Fem	-0.031*** (0.006)	-0.034*** (0.006)	-0.035*** (0.006)	-0.054 (0.047)	-0.055 (0.051)	-0.046 (0.047)	-0.044*** (0.011)	-0.035*** (0.007)
Fem*Math Teacher Stereotypes		-0.007 (0.007)	-0.004 (0.007)	-0.007 (0.007)	-0.007 (0.007)	-0.005 (0.007)	-0.013 (0.009)	
Fem*Lit Teacher Stereotypes							0.002 (0.011)	-0.001 (0.007)
Constant	0.196*** (0.003)	0.196*** (0.003)	0.124*** (0.006)	0.106*** (0.007)	0.109*** (0.008)	0.080*** (0.008)	0.198*** (0.004)	0.203*** (0.003)
Mean Y Fem	0.165	0.165	0.090	0.090	0.091	0.091	0.160	0.168
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R <sup>2</sup>	0.235	0.235	0.278	0.280	0.283	0.402	0.201	0.219
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	No	Yes	No	No

Notes: This table reports OLS estimates of equation 1, where the dependent variable is the high-school track recommendation of teachers. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. "Sq. Math Test 8" is the second order polynomial of test score in grade 8. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table X. Estimation of the effect of teachers' gender bias on self-confidence - class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)</b>								
Female	-0.092*** (0.029)	-0.078*** (0.027)	-0.077*** (0.028)	-0.099 (0.064)	-0.076 (0.063)	0.187 (0.214)	-0.089** (0.042)	-0.042 (0.029)
Fem*Math Teacher Stereotypes		-0.048** (0.024)	-0.048* (0.025)	-0.055** (0.025)	-0.042* (0.024)	-0.059* (0.032)	-0.065** (0.030)	
Fem*Lit Teacher Stereotypes							-0.005 (0.040)	-0.041 (0.028)
Constant	0.840*** (0.015)	0.838*** (0.018)	0.830*** (0.033)	0.843*** (0.049)	0.868*** (0.047)	0.866*** (0.047)	0.832*** (0.021)	0.838*** (0.018)
Sq. Std Test score math 6	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sq. Std Test score math 8	No	No	No	No	Yes	Yes	No	No
Obs.	789	789	789	789	789	789	461	717
R <sup>2</sup>	0.153	0.258	0.259	0.271	0.313	0.327	0.278	0.269
<b>Panel B- Dependent Variable: Being good/mediocre at reading (vs. being bad)</b>								
Female	0.041* (0.023)	0.050** (0.023)	0.048** (0.023)	0.056 (0.044)	0.051 (0.044)	0.035 (0.288)	0.062*** (0.023)	0.028 (0.018)
Fem*Math Teacher Stereotypes		0.028 (0.018)	0.028 (0.019)	0.030 (0.020)	0.032 (0.020)	0.028 (0.020)	0.043** (0.021)	
Fem*Lit Teachers Stereotypes							-0.047* (0.024)	-0.033* (0.017)
Constant	0.922*** (0.012)	0.914*** (0.015)	0.930*** (0.022)	0.928*** (0.032)	0.943*** (0.032)	0.945*** (0.034)	0.926*** (0.017)	0.934*** (0.013)
Sq. Test score read 6	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sq. Test score read 8	No	No	No	No	Yes	Yes	No	No
Obs.	705	705	705	705	705	705	418	637
R <sup>2</sup>	0.205	0.216	0.219	0.223	0.234	0.251	0.139	0.164
<b>Panel C- Dependent Variable: Average own ability in other subjects</b>								
Female	0.029 (0.026)	0.012 (0.026)	0.014 (0.025)	0.010 (0.054)	0.003 (0.053)	-0.313 (0.205)	0.012 (0.042)	0.020 (0.028)
Fem*Math Teacher Stereotypes		-0.013 (0.023)	-0.013 (0.023)	-0.018 (0.025)	-0.023 (0.024)	-0.035 (0.028)	-0.010 (0.031)	
Fem*Lit Teacher Stereotypes							-0.003 (0.040)	-0.009 (0.029)
Constant	1.672*** (0.013)	1.667*** (0.018)	1.654*** (0.024)	1.656*** (0.037)	1.658*** (0.039)	1.669*** (0.039)	1.672*** (0.020)	1.678*** (0.015)
Sq. Std Test score math 6	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sq. Std Test score math 8	No	No	No	No	Yes	Yes	No	No
Obs.	852	852	852	852	852	852	499	776
R <sup>2</sup>	0.147	0.178	0.185	0.189	0.198	0.220	0.141	0.149
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	No	No	Yes	No	No

Notes: This table reports OLS estimates of equation 1, where the dependent variable is self-stereotypes in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

## Online Appendix

# Implicit Stereotypes: Evidence from Teachers' Gender Bias

Michela Carlana\*

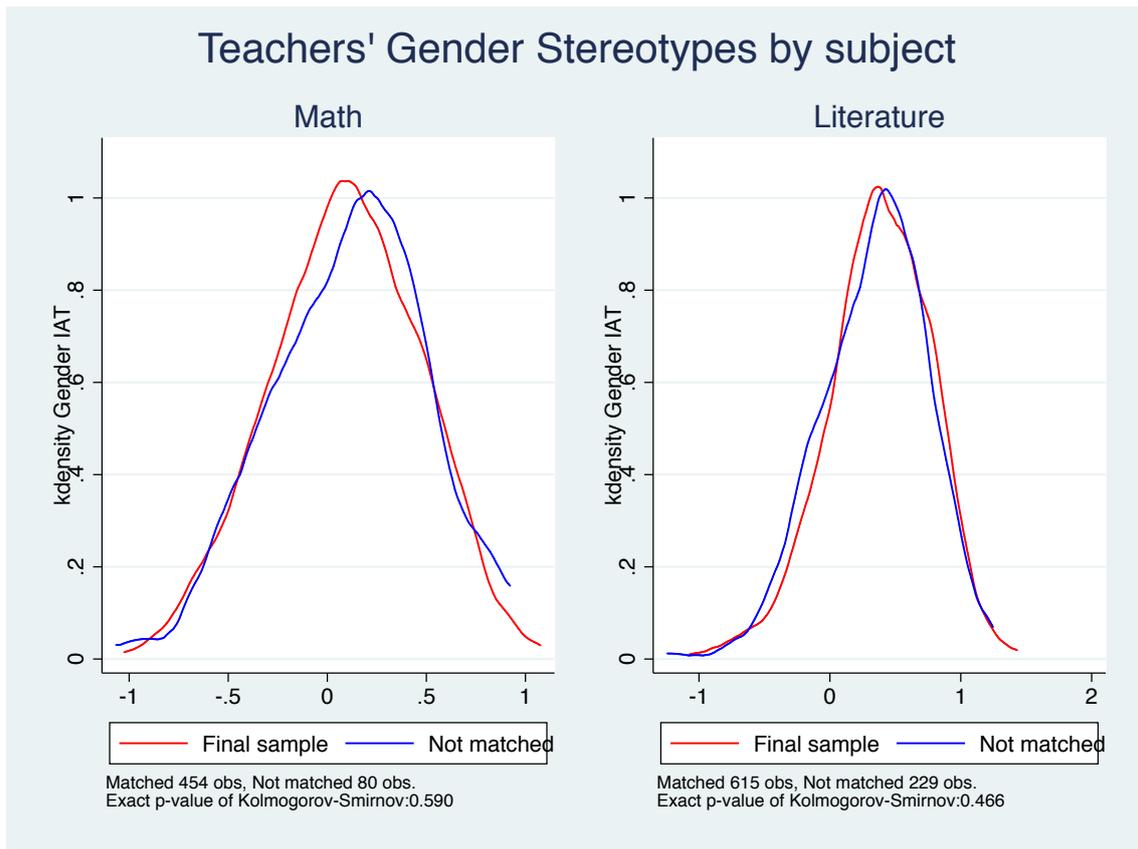
January 2019

---

\*Harvard Kennedy School (e-mail: [michela\\_carlana@hks.harvard.edu](mailto:michela_carlana@hks.harvard.edu)).

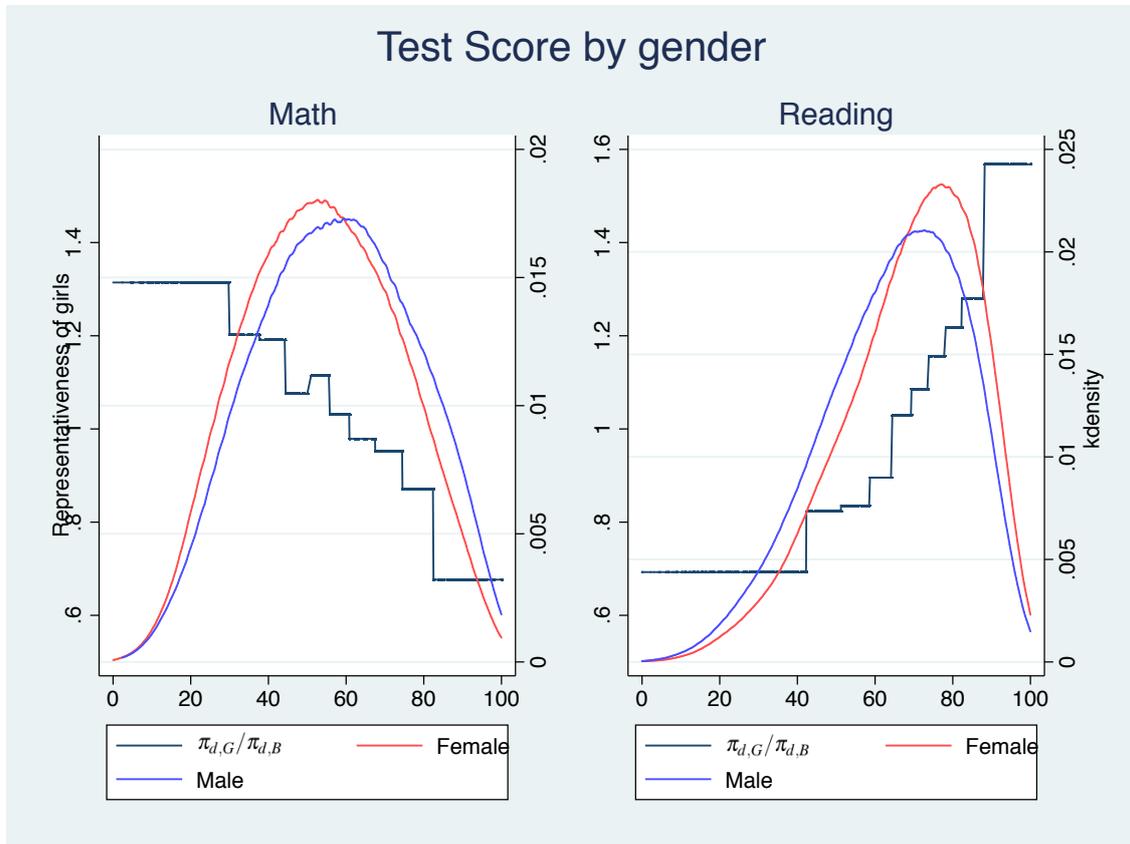
## A Additional Figures and Tables

Figure A.I. Teachers' Implicit Gender Stereotypes (IAT measure) by subject of matched and unmatched sample



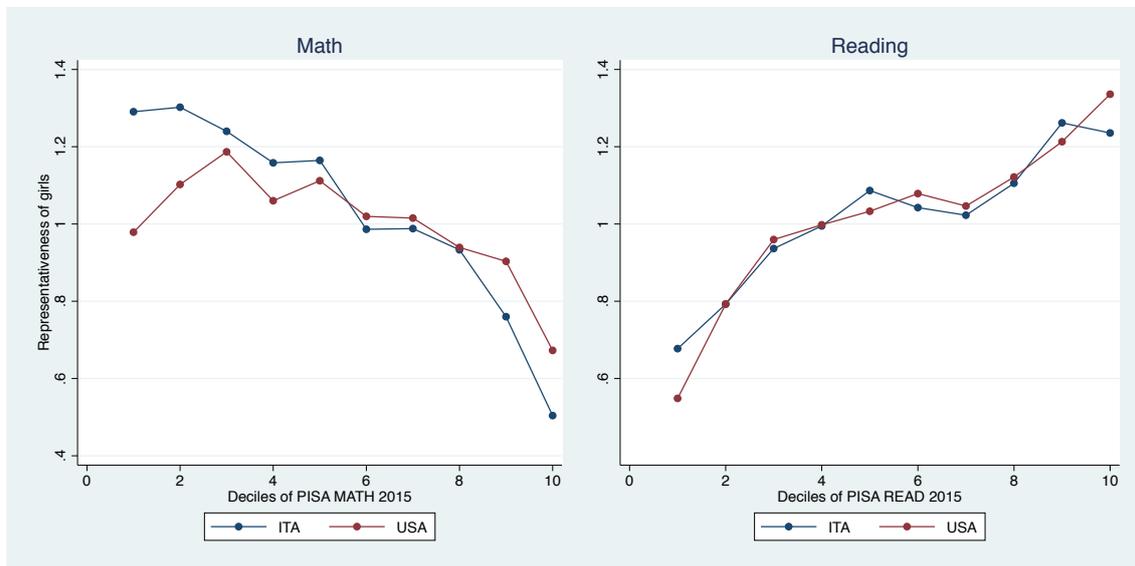
*Notes:* This graph shows the distribution of Gender-Science IAT scores for math and literature teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes. The graph provides evidence that teachers used in this paper are similar in terms of gender stereotypes to those who completed the survey but are not included in the main analysis because I do not have outcomes of their students.

Figure A.II. Gender differences in math and reading standardized test score



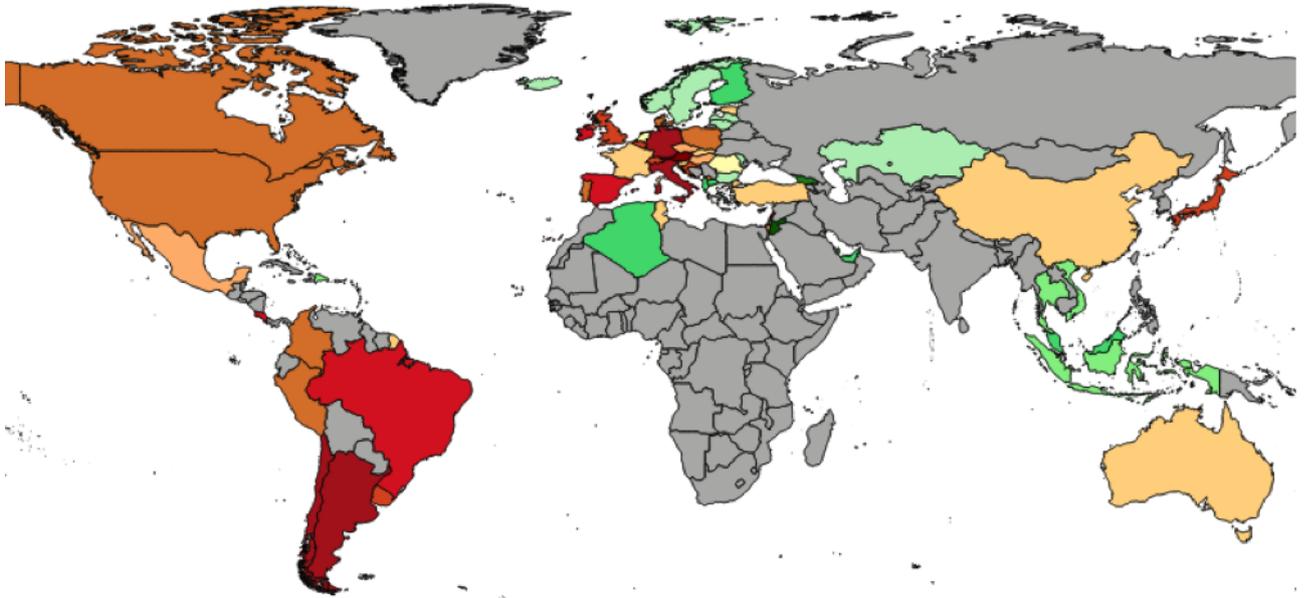
*Notes:* This graph shows the difference in math and reading test scores in grade 8 between females and males in the sample of Italian middle schools.

Figure A.III. Representativeness of girls with respect to boys in PISA 2015 in US and Italy



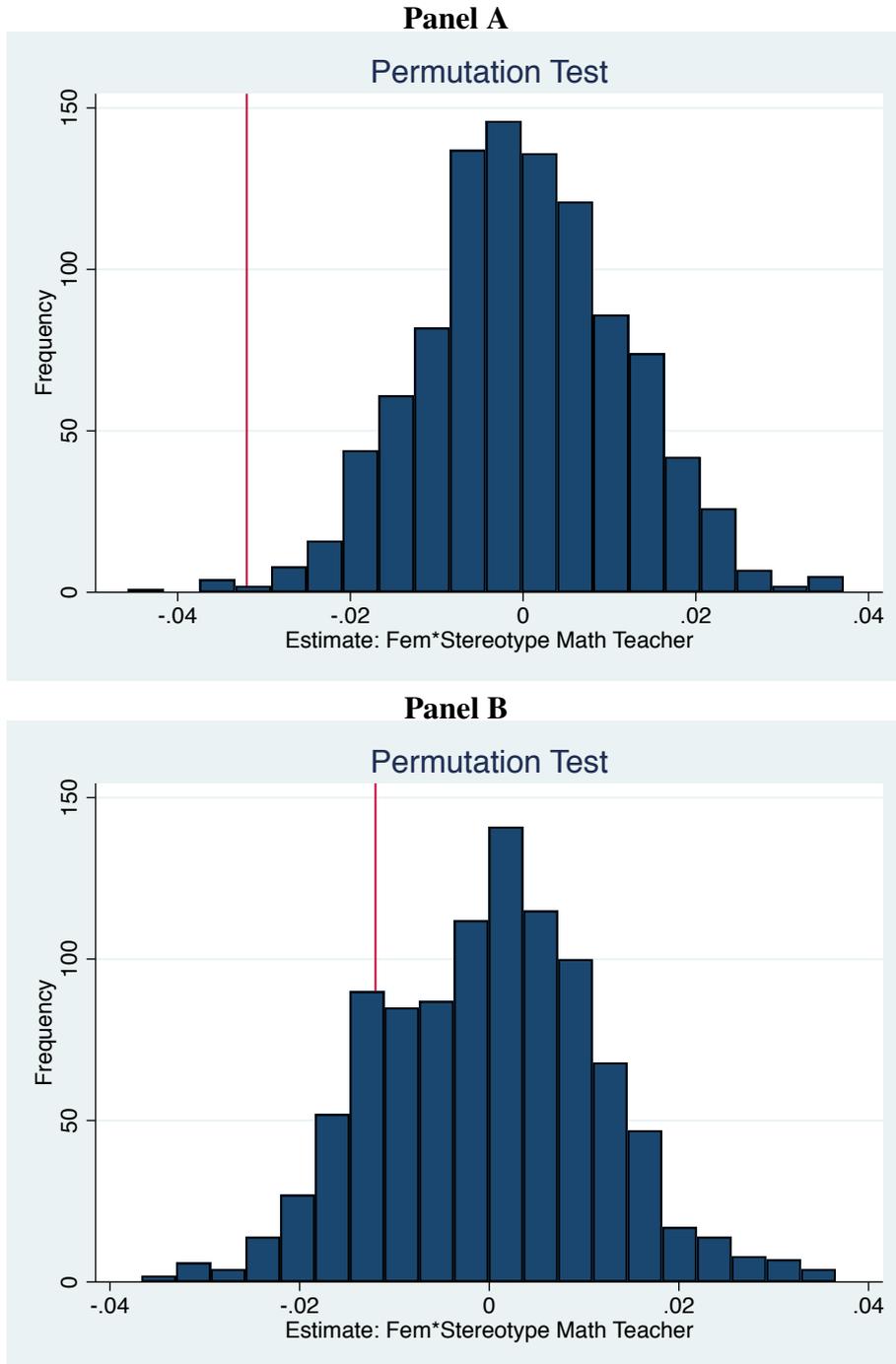
Notes: This graph shows the representativeness of girls with respect to boys in PISA 2015 (math and reading).

Figure A.IV. Gender differences in math standardized test score PISA



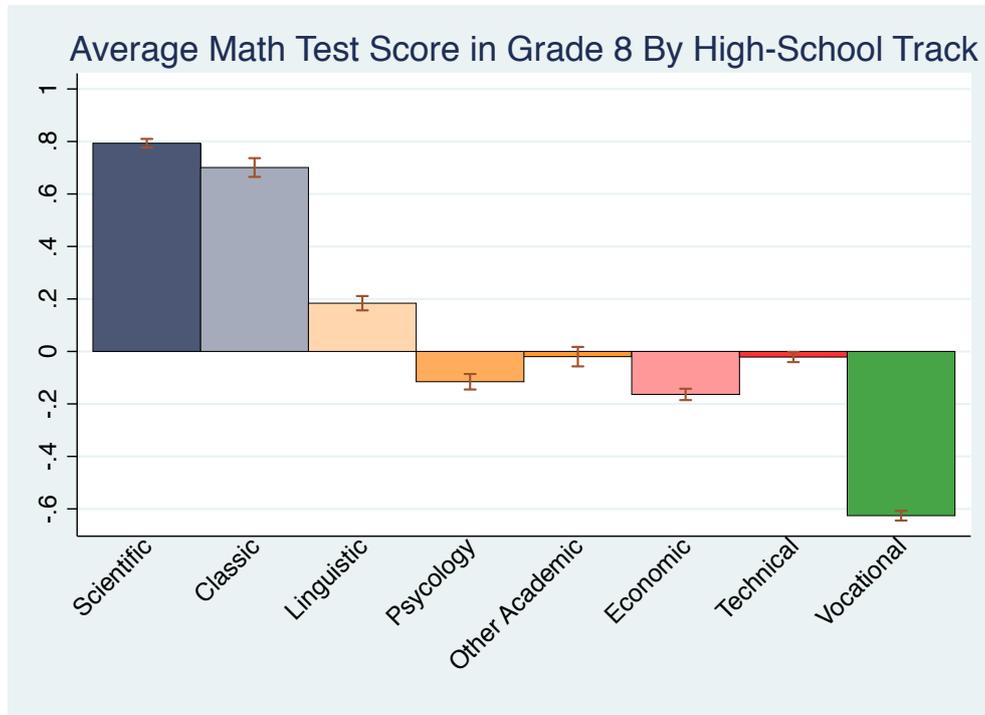
*Notes:* This graph shows the difference in math PISA test scores between females and males. Countries in which girls lag behind are colored in red, while countries where boys lag behind are colored in green.  
*Source:* Author's calculation on PISA data (2015).

Figure A.V. Permutation Tests



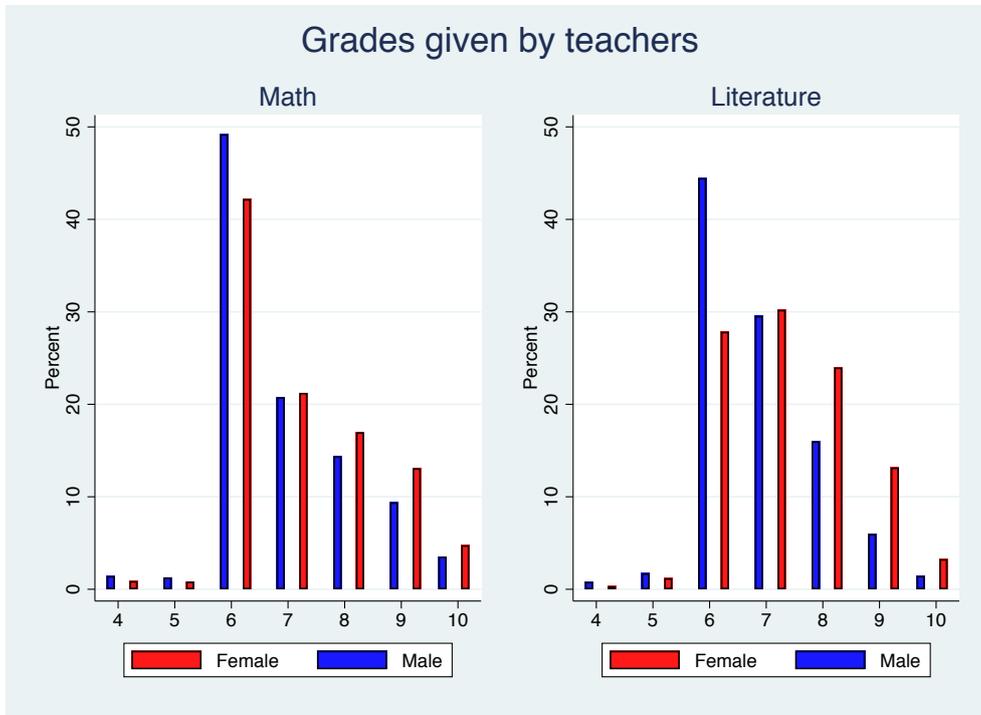
*Notes:* This figures plots the “Fem\*Stereotypes” of a permutation test that runs the main regression in equation (1) 1000 times randomly assigning the stereotypes to math teachers in Panel A and literature teachers in Panel B. The vertical (red) line signifies the coefficient from the main regression in column 2 of Table V.

Figure A.VI. Average Standardized Test Score in Grade 8 by High-School Track



Notes: This graph shows the average standardized test score in grade 8 by high-school track selected.

Figure A.VII. Grades given by teachers



Notes: This graph shows the distribution of grades given by teachers at the end of the school year.

Table A.I. Balance Table: students in the sample vs. other students

<b>Panel A: Comparison between students in the sample and other Italian students</b>				
Variable	(1) Other Italian Students	(2) Sample	(3) Diff.	(4) Norm. Diff.
Female	0.410 (0.492)	0.455 (0.498)	0.002 (0.003)	0.064
Immigrant	0.096 (0.295)	0.217 (0.412)	0.118 (0.011)***	0.238
Test score grade 8	56.636 (19.037)	55.547 (19.699)	-0.824 (0.696)	-0.040
Mother: Less than Diploma	0.244 (0.430)	0.227 (0.419)	-0.047 (0.016)***	-0.029
Mother: Diploma	0.330 (0.470)	0.378 (0.485)	0.014 (0.015)	0.071
Mother: More than Diploma	0.095 (0.294)	0.138 (0.344)	0.035 (0.017)**	0.093
Father: Less than Diploma	0.282 (0.450)	0.263 (0.440)	-0.053 (0.018)***	-0.031
Father: Diploma	0.291 (0.454)	0.334 (0.472)	0.013 (0.013)	0.065
Father: More than Diploma	0.084 (0.277)	0.126 (0.332)	0.037 (0.017)**	0.098
Mother: Low Occupation	0.384 (0.486)	0.376 (0.484)	-0.052 (0.019)***	-0.011
Mother: Intermediate Occupation	0.220 (0.414)	0.289 (0.453)	0.050 (0.012)***	0.113
Mother: High Occupation	0.071 (0.257)	0.094 (0.292)	0.017 (0.011)	0.059
Father: Low Occupation	0.242 (0.428)	0.282 (0.450)	0.017 (0.017)	0.065
Father: Intermediate Occupation	0.267 (0.442)	0.291 (0.454)	-0.003 (0.012)	0.039
Father: High Occupation	0.143 (0.350)	0.162 (0.368)	0.004 (0.017)	0.037
Class size	22.151 (3.829)	22.526 (2.668)	0.379 (0.173)**	0.080
Observations	3,100,239	41,733	3,141,972	

Table A.I. Balance Table: students in the sample vs. other students (*cont.*)

<b>Panel B: Comparison between students in the sample and other students in the same provinces</b>				
Variable	(1) Other Italian Students	(2) Sample	(3) Diff.	(4) Norm. Diff.
Female	0.407 (0.491)	0.455 (0.498)	0.003 (0.003)	0.069
Immigrant	0.134 (0.341)	0.217 (0.412)	0.081 (0.011)***	0.155
Test score grade 8	56.563 (19.039)	55.547 (19.699)	-1.177 (0.700)*	-0.037
Mother: Less than Diploma	0.194 (0.396)	0.227 (0.419)	0.012 (0.017)	0.056
Mother: Diploma	0.362 (0.481)	0.378 (0.485)	-0.026 (0.016)*	0.023
Mother: More than Diploma	0.119 (0.324)	0.138 (0.344)	0.006 (0.017)	0.039
Father: Less than Diploma	0.238 (0.426)	0.263 (0.440)	-0.002 (0.018)	0.040
Father: Diploma	0.317 (0.465)	0.334 (0.472)	-0.020 (0.014)	0.025
Father: More than Diploma	0.107 (0.309)	0.126 (0.332)	0.008 (0.017)	0.042
Mother: Low Occupation	0.319 (0.466)	0.376 (0.484)	0.023 (0.019)	0.085
Mother: Intermediate Occupation	0.273 (0.446)	0.289 (0.453)	-0.015 (0.013)	0.026
Mother: High Occupation	0.094 (0.292)	0.094 (0.292)	-0.012 (0.011)	-0.001
Father: Low Occupation	0.225 (0.417)	0.282 (0.450)	0.035 (0.017)**	0.093
Father: Intermediate Occupation	0.276 (0.447)	0.291 (0.454)	-0.017 (0.012)	0.023
Father: High Occupation	0.168 (0.374)	0.162 (0.368)	-0.028 (0.018)	-0.013
Class size	22.537 (3.162)	22.526 (2.668)	0.015 (0.180)	-0.003
Observations	418,408	41,733	460,141	

*Notes:* data from the standardized test score INVALSI of all Italian students in grade 8 from 2011-12 to 2016-17. Students in the sample are those for which we have information on the standardized test scores in mathematics and the IAT of their math teacher and/or the standardized test scores in reading and the IAT of their literature teacher. The normalized difference shown in column 4 is the formula recommended by [Imbens and Wooldridge \(2009\)](#). More details are reported in footnote 11 of the paper.

Table A.II. Balance table of the differences between teachers matched (not matched) with students graduating from 2012 to 2017

	Math Teachers				Literature Teachers			
	Not matched	Matched	Dif.	se	Not matched	Matched	Dif.	se
Female	0.771	0.811	-0.039	(0.047)	0.845	0.899	-0.055*	(0.024)
Born in the North	0.573	0.611	-0.038	(0.061)	0.547	0.719	-0.172***	(0.037)
Age	42.158	48.806	-6.648***	(1.182)	40.858	49.550	-8.692***	(0.688)
Full time contract	0.461	0.788	-0.327***	(0.053)	0.643	0.944	-0.302***	(0.025)
Yeas of experience	10.455	18.829	-8.375***	(1.457)	11.472	21.668	-10.196***	(0.795)
Teaching in 2015-16	0.415	0.869	-0.455***	(0.044)	0.444	0.952	-0.508***	(0.025)
Children	0.590	0.683	-0.092	(0.056)	0.504	0.706	-0.201***	(0.036)
Number of children	1.659	1.876	-0.217	(0.136)	1.573	1.781	-0.208*	(0.090)
Number of daughters	0.864	0.866	-0.002	(0.125)	0.909	0.845	0.064	(0.083)
Low edu Mother	0.452	0.518	-0.066	(0.063)	0.346	0.492	-0.145***	(0.040)
Middle edu Mother	0.384	0.331	0.053	(0.060)	0.434	0.373	0.061	(0.040)
High edu Mother	0.164	0.151	0.013	(0.046)	0.220	0.135	0.084**	(0.030)
Advanced STEM	0.282	0.231	0.051	(0.052)	0.000	0.000	0.000	(0.000)
Math Olympiad	0.051	0.163	-0.112*	(0.043)	0.000	0.000	0.000	(0.000)
Refresher courses	0.821	0.914	-0.094*	(0.037)	0.825	0.926	-0.101***	(0.023)
Degree Laude	0.250	0.209	0.041	(0.055)	0.372	0.326	0.046	(0.040)
IAT Gender	0.116	0.091	0.025	(0.046)	0.338	0.377	-0.039	(0.030)
IAT Race	0.461	0.465	-0.004	(0.031)	0.433	0.466	-0.033	(0.020)
Boys better in Invalsi	0.245	0.213	0.031	(0.063)	0.088	0.100	-0.011	(0.031)
Girls better in Invalsi	0.367	0.307	0.060	(0.071)	0.531	0.537	-0.006	(0.053)
Gender Equal in Invalsi	0.388	0.480	-0.092	(0.076)	0.381	0.363	0.018	(0.051)
Satisfy with teacher job	3.718	3.717	0.001	(0.109)	3.800	3.905	-0.105	(0.068)
WVS Gender Equality	0.167	0.158	0.009	(0.045)	0.112	0.105	0.006	(0.024)
Gender Dif Innate Ability	1.653	1.512	0.141	(0.095)	1.475	1.413	0.062	(0.055)
Observations	83		454		238	615		

Notes: First hand data from teachers' questionnaire. We compare teachers matched with students' data with those not matched. Notice that for 3 math and 9 literature teachers, there is no IAT test because they did not complete the test.

Table A.III. The impact of teacher characteristics on students' improvement in performance

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Dependent variable: Std Math test score in grade 8</b>						
<b>Panel A: X=</b>	<b>Teacher Female</b>		<b>Math Teachers Advanced STEM</b>		<b>Degree cum laude</b>	
X	0.154*** (0.051)	0.079** (0.033)	-0.030 (0.058)	-0.063** (0.031)	0.048 (0.048)	0.052* (0.027)
Fem*X	0.026 (0.033)	0.026 (0.029)	-0.020 (0.035)	-0.023 (0.032)	0.019 (0.028)	0.015 (0.025)
Obs.	30359	30359	30359	30359	30359	30359
R <sup>2</sup>	0.012	0.178	0.009	0.178	0.009	0.178
<b>Dependent variable: Std Math test score in grade 8</b>						
<b>Panel B: X=</b>	<b>Math Olympics</b>		<b>Math Teachers Full-time contract</b>		<b>High experience</b>	
X	0.050 (0.055)	0.074** (0.033)	0.195*** (0.048)	0.159*** (0.036)	0.236*** (0.064)	0.135** (0.061)
Fem*X	-0.013 (0.034)	-0.024 (0.030)	-0.027 (0.040)	-0.030 (0.033)	-0.056 (0.058)	-0.079 (0.050)
Obs.	30359	30359	30359	30359	30359	30359
R <sup>2</sup>	0.009	0.178	0.012	0.179	0.012	0.178
<b>Dependent variable: Std Reading test score in grade 8</b>						
<b>Panel C: X=</b>	<b>Teacher Female</b>		<b>Literature Teachers Full-time contract</b>		<b>High experience</b>	
X	0.103** (0.051)	0.076*** (0.029)	0.082 (0.081)	0.008 (0.059)	0.053 (0.054)	0.064* (0.037)
Fem*X	-0.048 (0.040)	-0.040 (0.035)	-0.024 (0.072)	-0.018 (0.056)	0.008 (0.048)	-0.025 (0.041)
Obs.	29486	29486	29486	29486	29486	29486
R <sup>2</sup>	0.014	0.213	0.014	0.213	0.015	0.213
Student controls	No	Yes	No	Yes	No	Yes
School FE	No	Yes	No	Yes	No	Yes

*Notes:* This table reports OLS estimates of the impact of math teachers' characteristics on improvements in math performance of their students. The unit of observation is the student. Standard errors (in parentheses) are robust and clustered at the teacher level. All columns include a dummy for the gender of the student ("Fem"). "High experience" means more than 15 years, "Mid experience" from 5 to 15. Student controls include parents' education and occupation and immigration status interacted with student gender. An intermediate level of experience and a missing category for experience (and the interactions with female) are included in columns 5-6 of Panel B and C. The level of this variable is statistically significant only for column 5 of Panel B. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.IV. Correlation between teachers' characteristics and Gender IAT Score

	<b>Math Teachers</b>		<b>Literature Teachers</b>	
	<b>Matched sample</b>	<b>All teachers</b>	<b>Matched sample</b>	<b>All teachers</b>
	(1)	(2)	(3)	(4)
Female	-0.166*** (0.053)	-0.184*** (0.045)	0.552*** (0.057)	0.568*** (0.047)
Born in the North	-0.055 (0.038)	-0.066* (0.034)	-0.054* (0.030)	-0.041 (0.025)
Age	-0.014 (0.025)	-0.028 (0.021)	0.000 (0.018)	-0.001 (0.015)
Age sq.	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
High Edu Mother	0.031 (0.037)	0.013 (0.036)	-0.006 (0.033)	0.008 (0.027)
Children	-0.003 (0.059)	0.024 (0.055)	-0.052 (0.046)	-0.046 (0.039)
Daughters	0.054 (0.051)	0.017 (0.050)	0.092** (0.039)	0.054 (0.033)
Degree Laude	-0.093** (0.041)	-0.060 (0.042)	-0.000 (0.029)	0.028 (0.026)
Full time contract	-0.056 (0.071)	-0.004 (0.057)	0.017 (0.067)	0.056 (0.049)
High Experience	-0.030 (0.100)	0.020 (0.086)	-0.098 (0.112)	-0.065 (0.072)
Med Experience	0.052 (0.080)	0.087 (0.064)	-0.002 (0.097)	0.021 (0.064)
WVS Gender Equality	-0.025 (0.052)	-0.009 (0.048)	0.041 (0.045)	0.053 (0.035)
Gender Dif Innate Ability	0.031 (0.024)	0.027 (0.020)	-0.014 (0.024)	-0.001 (0.020)
Advanced STEM	-0.094** (0.044)	-0.093** (0.042)		
Math Olympiad	0.081 (0.057)	0.074 (0.054)		
Obs.	454	534	615	844
R <sup>2</sup>	0.105	0.101	0.223	0.249

*Notes:* This table reports OLS estimates of the correlation between literature teacher stereotypes measured by IAT score and own teacher characteristics. The unit of observation is teacher  $t$  in school  $s$ . Standard errors (in parentheses) are robust and clustered at the school level. The description of all variables is the same as in Table III. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. The restricted sample includes data of teachers used in the main regressions of this paper. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.V. Exogeneity of assignment of students to math teachers with different bias

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A- Dependent Variable: Math Teacher implicit gender bias (standardized)</b>						
Share Fem	0.055 (0.934)	0.048 (0.925)	0.004 (0.942)	-0.030 (0.904)	-0.125 (0.871)	0.589 (1.350)
Share Male HighEduMother		-1.209** (0.563)			-1.203* (0.638)	-1.203 (0.949)
Share Fem HighEduMother		0.255 (0.448)			0.202 (0.549)	-0.665 (0.902)
Share Male MedOccFather			-0.074 (0.772)		0.428 (0.840)	-0.229 (1.037)
Share Fem MedOccFather			-0.026 (0.623)		-0.187 (0.796)	0.413 (1.098)
Share Male HighOccFather			0.010 (0.978)		0.436 (0.965)	1.496 (1.355)
Share Fem HighOccFather			-0.830 (0.834)		-0.817 (1.002)	0.031 (1.615)
Share Male Immigrant				0.447 (0.618)	0.574 (0.631)	0.912 (0.927)
Share Fem Immigrant				-0.464 (0.620)	-0.758 (0.710)	-0.703 (0.864)
Male Average Std Math 5						-0.084 (0.259)
Fem Average Std Math 5						-0.180 (0.212)
Obs.	454	454	454	454	454	282
R <sup>2</sup>	0.202	0.215	0.207	0.205	0.226	0.239

Table A.V. Exogeneity of assignment of students to math teachers with different bias

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel B- Dependent Variable: Literature Teacher implicit gender bias (standardized)</b>						
Share Fem	-0.367 (0.594)	-0.387 (0.607)	-0.328 (0.603)	-0.395 (0.610)	-0.384 (0.626)	0.009 (0.843)
Share Male HighEduMother		0.005 (0.416)			0.064 (0.429)	0.077 (0.653)
Share Fem HighEduMother		-0.306 (0.335)			-0.359 (0.433)	-0.972* (0.509)
Share Male MedOccFather			-0.889* (0.462)		-0.981** (0.491)	-1.139 (0.759)
Share Fem MedOccFather			0.052 (0.425)		0.083 (0.449)	0.575 (0.597)
Share Male HighOccFather			-0.095 (0.555)		-0.208 (0.594)	0.359 (0.822)
Share Fem HighOccFather			0.031 (0.527)		0.205 (0.622)	0.271 (0.994)
Share Male Immigrant				-0.394 (0.424)	-0.497 (0.452)	-0.874 (0.551)
Share Fem Immigrant				0.009 (0.363)	-0.028 (0.391)	0.104 (0.523)
Male Average Std Math 5						-0.126 (0.161)
Fem Average Std Math 5						0.090 (0.133)
Obs.	616	616	616	616	616	357
R <sup>2</sup>	0.177	0.181	0.184	0.179	0.191	0.233

*Notes:* This table reports OLS estimates of the correlation between teacher stereotypes measured by IAT score and student characteristics. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the school level. The number of clusters is 98 in columns 1-6 and 73 in column 7. All regressions include controls for the order of IAT in the questionnaire administered. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VI. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

	All		Female Teachers		Male Teachers	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Dep. Variable - Std Math Grade 8</b>						
Female	-0.198*** (0.013)	-0.033 (0.082)	-0.196*** (0.015)	0.002 (0.094)	-0.203*** (0.029)	-0.193 (0.202)
Fem*Teacher Stereotypes	-0.032** (0.013)	-0.031*** (0.011)	-0.030** (0.014)	-0.031** (0.012)	-0.040 (0.029)	-0.059** (0.028)
Obs.	30359	30359	25550	25550	4809	4809
R <sup>2</sup>	0.206	0.276	0.205	0.275	0.193	0.269
<b>Panel B: Dep. Variable - Std Reading Grade 8</b>						
Female	0.225*** (0.013)	0.308*** (0.093)	0.220*** (0.014)	0.266*** (0.094)	0.258*** (0.050)	0.366 (0.265)
Fem*Teacher Stereotypes	-0.012 (0.013)	-0.001 (0.013)	-0.007 (0.015)	0.001 (0.014)	-0.011 (0.039)	-0.014 (0.046)
Obs.	29484	29484	26901	26901	2583	2583
R <sup>2</sup>	0.181	0.291	0.181	0.291	0.179	0.291
Student controls	No	Yes	No	Yes	No	Yes
Teacher controls	No	Yes	No	Yes	No	Yes

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math and reading standardized test score in grade 8 in Panel A and B, respectively. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VII. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

	<b>Dep. Variable: Std Math 8th grade</b>					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.154*** (0.018)	-0.131*** (0.028)	0.005 (0.084)	-0.148*** (0.021)	-0.122*** (0.030)	0.014 (0.085)
Fem *No Bias				-0.051 (0.032)	-0.054* (0.030)	-0.051* (0.029)
Fem *Male-Math				-0.044 (0.027)	-0.049* (0.026)	-0.047* (0.025)
Fem *IAT>0 Male-Math	-0.049** (0.024)	-0.048** (0.022)	-0.048** (0.022)			
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Obs.	30359	30359	30359	30359	30359	30359
$R^2$	0.206	0.275	0.276	0.206	0.275	0.276

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Instead of using a continuous variable as teacher bias, I use categorical variables. In columns 1-3, we consider a positive or negative sign in the IAT score. In columns 4-6, we consider the thresholds defined by [Greenwald et al. \(2003\)](#) where no bias is the interval of IAT raw score between -0.15 and +0.15. Standard errors (in parentheses) are robust and clustered at the teacher level. The number of clusters is 454. The variable “Fem” indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher’s mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VIII. Estimation of the effect of teachers' gender stereotypes

Dependent Variable: Math standardized test score in grade 8						
Heterogeneous effects by	Student Characteristics				Interaction time with teacher	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.032 (0.082)	0.089 (0.120)	-0.026 (0.082)	-0.032 (0.082)	-0.037 (0.082)	-0.015 (0.085)
Fem*Teacher Stereotypes(IAT)	-0.031*** (0.011)	-0.096*** (0.027)	-0.019 (0.020)	-0.030** (0.013)	-0.029** (0.012)	-0.037*** (0.013)
Fem*IAT*Top tercile Math6		0.136*** (0.043)				
Fem*IAT*Middle tercile Math6		0.024 (0.039)				
Fem*IAT*HighEduM			-0.000 (0.026)			
Fem*IAT*Missing EduM			-0.051 (0.031)			
Fem*IAT*Immigrant				-0.008 (0.030)		
Fem*IAT*Extended School Day					-0.009 (0.029)	
Fem*IAT*New Math Teacher						0.031 (0.031)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	30359	9540	30359	30359	30359	30359
R <sup>2</sup>	0.276	0.575	0.276	0.276	0.276	0.276

*Notes:* This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on math standardized test score in grade 8 by observable characteristics of the student and by interaction time with teacher. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The number of clusters is 454. The number of observations is lower in the last column because for the first two cohorts of students who started middle school before 2011, I do not have information on who their teacher was in grade 6 and/or 7. The variable "Fem" indicates the gender of the student, "HighEduM" whether the mother has at least a diploma, "Missing EduM" whether the information on education of the mother is missing, "tercile Math6" is the tercile of standardized test score in math in grade 6, and "Immigrant" indicates whether the student is not Italian citizen. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. Regressions are all fully saturated even if not all interactions are shown in the table. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.IX. Estimation of the effect of teachers' gender stereotypes for one year on math standardized test score in grade 6 and 8 - class FE regression

	Std Math 6th grade			Std Math 8th grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.209*** (0.029)	-0.202*** (0.027)	0.140 (0.227)	-0.190*** (0.032)	-0.192*** (0.030)	0.033 (0.171)
Fem*Math Teacher Stereotypes	0.004 (0.033)	0.005 (0.030)	0.010 (0.030)	0.011 (0.028)	-0.006 (0.027)	0.010 (0.029)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Student Controls*Fem	No	No	Yes	No	No	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Obs.	5532	5532	5532	4695	4695	4695
$R^2$	0.214	0.275	0.278	0.227	0.298	0.302

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 6 in columns 1-3 and grade 8 in columns 4-6 of school  $s$ . The sample is restricted to teachers teaching in grade 6 for columns 1-3 and for teachers assigned to students only in grade 8 for columns 4-6. Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.X. Estimation of the effect of teachers' gender stereotypes on the probability of missing standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)
<b>Panel A- Dependent Variable: Missing math standardized test score in grade 8</b>					
Female	-0.025*** (0.003)	-0.023*** (0.003)	-0.014*** (0.003)	-0.006* (0.004)	0.023 (0.017)
Fem*Math Teacher Stereotypes		0.005 (0.003)	0.003 (0.003)	0.002 (0.003)	0.002 (0.003)
Fem* Teacher Fem					0.010 (0.006)
Fem* Teacher. Born North					-0.004 (0.005)
Fem*Advanced STEM Teacher					-0.005 (0.006)
Obs.	32642	32642	32642	32642	32642
R <sup>2</sup>	0.102	0.102	0.377	0.380	0.380
<b>Panel B- Dependent Variable: Missing reading standardized test score in grade 8</b>					
Female	-0.024*** (0.003)	-0.023*** (0.003)	-0.012*** (0.003)	-0.004 (0.004)	-0.004 (0.021)
Fem*Lit Teacher Stereotypes		-0.004 (0.003)	-0.002 (0.003)	-0.003 (0.003)	-0.003 (0.003)
Fem* Teacher Fem					0.005 (0.010)
Fem* Teacher. Born North					-0.011* (0.006)
Obs.	31701	31701	31701	31701	31701
R <sup>2</sup>	0.089	0.089	0.367	0.368	0.368
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is the probability of not taking the math or reading standardized test score in grade 8 in Panel A and B, respectively. The data may be missing because the student did not take the test or because the school did not correctly match the test score with the administrative registry. The dependent variables refer to math teachers in Panel A and to literature teachers in Panel B. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level (454 in Panel A and 615 in Panel B). The variable “Fem” indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract and education of the teacher’s mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XI. Estimation of the effect of teachers' gender stereotypes on standardized test score in math in grade 8 for the different cohorts

<b>Dependent Variable: Math standardized test score in grade 8</b>			
	(1)	(2)	(3)
	All	2012-2016	2017
Graduation			
Female	-0.032 (0.082)	-0.026 (0.101)	-0.030 (0.142)
Fem*Teachers' Stereotypes	-0.031*** (0.011)	-0.029** (0.014)	-0.037* (0.021)
Class FE	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes
Obs.	30359	22657	7702
$R^2$	0.276	0.278	0.275

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XII. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression -sample restricted to cohorts by schools where students are randomly assigned according to all characteristics by gender

	(1)	(2)	(3)	(4)	(5)
<b>Panel A: Dependent Variable - Math standardized test score in grade 8</b>					
Female	-0.175*** (0.016)	-0.189*** (0.019)	-0.190*** (0.018)	-0.158*** (0.035)	-0.116 (0.109)
Fem*Math Teacher Stereotypes		-0.031* (0.018)	-0.033* (0.017)	-0.033* (0.017)	-0.035** (0.018)
Fem*Math Teacher Fem					0.058 (0.041)
Fem*North Math Teacher					0.007 (0.032)
Fem*Advanced STEM Teacher					0.004 (0.035)
Obs.	15781	15781	15781	15781	15781
R <sup>2</sup>	0.201	0.201	0.272	0.273	0.273
<b>Panel B: Dependent Variable - Reading standardized test score in grade 8</b>					
Female	0.226*** (0.016)	0.230*** (0.017)	0.228*** (0.016)	0.245*** (0.033)	0.430*** (0.135)
Fem*Literature Teacher Stereotypes		-0.012 (0.017)	-0.010 (0.016)	-0.011 (0.016)	-0.009 (0.016)
Fem*Literature Teacher Fem					-0.032 (0.059)
Fem*North Literature Teacher					0.002 (0.039)
Obs.	15916	15916	15916	15916	15916
R <sup>2</sup>	0.172	0.172	0.281	0.281	0.282
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. I restrict the sample to schools by cohorts where Pearson Chi-Square tests suggest statistically independence of all student characteristics (gender, education of the mother, occupation of the father, immigrant dummy, generation of immigration) and of all student characteristics by gender. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable “Fem” indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XIII. Estimation of the effect of teachers' gender stereotypes on math and reading standardized test score in grade 8 - class FE regression

Dependent Variable:	Std math 8th grade				Std reading 8th grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fem	-0.176*** (0.015)	-0.179*** (0.017)	-0.193*** (0.018)	-0.204 (0.134)	0.219*** (0.015)	0.214*** (0.017)	0.222*** (0.018)	0.392*** (0.105)
Fem*Lit Teach Stereotypes		0.008 (0.017)	0.006 (0.017)	0.022 (0.017)			-0.022 (0.017)	-0.016 (0.016)
Fem*Math Teach Stereotypes			-0.033** (0.016)	-0.029* (0.015)		-0.010 (0.017)	-0.011 (0.017)	-0.014 (0.016)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	No	No	No	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes
Obs.	18102	18102	18102	18102	18102	18102	18102	18102
R <sup>2</sup>	0.198	0.198	0.198	0.274	0.183	0.183	0.183	0.299

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math or reading standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . I restrict the sample to classes for which I have information on the implicit associations of both literature and math teachers. Standard errors (in parentheses) are robust and clustered at the class level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between students' gender and teacher characteristics. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XIV. Correlation between track choice and recommendation

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Dep Var</b>	<b>Vocational</b>			<b>Scientific</b>		
Vocational recommendation	0.327*** (0.010)	0.288*** (0.011)	0.280*** (0.011)			
Fem*Vocational recommendation	0.101*** (0.014)	0.094*** (0.014)	0.093*** (0.015)			
Scientific recommendation				0.670*** (0.011)	0.582*** (0.012)	0.561*** (0.013)
Fem*Scientific recommendation				-0.015 (0.015)	-0.006 (0.015)	0.020 (0.015)
Female	0.004 (0.003)	-0.003 (0.004)	0.016 (0.012)	-0.061*** (0.006)	-0.051*** (0.006)	-0.023** (0.010)
Constant	0.036*** (0.003)	0.043*** (0.004)	0.067*** (0.009)	0.162*** (0.004)	0.152*** (0.004)	0.096*** (0.008)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Sq. Test score 8	No	Yes	Yes	No	Yes	Yes
Student controls	No	No	Yes	No	No	Yes
Student controls*Fem	No	No	Yes	No	No	Yes
Obs.	20646	20646	20646	20646	20646	20646
R <sup>2</sup>	0.355	0.363	0.367	0.486	0.507	0.513

*Notes:* This table reports OLS estimates of the correlation between track choice and teachers' track recommendation, for vocational in columns 1-3 and scientific track in columns 4-6. The dataset used in this table is the same as in the main analysis of the paper, but here I restrict the sample to students for whom I have information on track choice, teachers' recommendation, and test score in grade 8. Standard errors (in parentheses) are robust and clustered at the class level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student in columns 3 and 6. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XV. Estimation of the effect of teachers' gender stereotypes on high-school track choice - school by cohort FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: Vocational High-School Track Choice</b>								
Fem	0.011** (0.006)	0.019*** (0.006)	0.019*** (0.006)	0.063 (0.046)	0.052 (0.045)	0.043 (0.043)	0.025*** (0.009)	0.009 (0.006)
Fem*Math Teach Stereotypes		0.018*** (0.006)	0.016** (0.006)	0.014** (0.006)	0.009 (0.006)	0.005 (0.006)	0.023** (0.009)	
Math Teach Stereotypes		-0.008* (0.005)	-0.008* (0.005)	-0.008 (0.005)	-0.008 (0.005)	-0.007 (0.005)	-0.002 (0.007)	
Fem*Literature Teach Stereotypes							0.010 (0.009)	-0.000 (0.006)
Literature Teach Stereotypes							-0.002 (0.006)	-0.000 (0.005)
Constant	0.163*** (0.004)	0.160*** (0.004)	0.186*** (0.008)	0.172*** (0.034)	0.198*** (0.035)	0.145*** (0.034)	0.153*** (0.006)	0.162*** (0.004)
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R <sup>2</sup>	0.067	0.068	0.116	0.120	0.108	0.176	0.078	0.068
Effect on girls		0.010** (0.005)	0.007 (0.005)	0.006 (0.005)	0.002 (0.005)	-0.002 (0.005)	0.021*** (0.006)	
<b>Panel B- Dependent Variable: Scientific High-School Track Choice</b>								
Fem	-0.089*** (0.007)	-0.093*** (0.007)	-0.091*** (0.007)	-0.029 (0.048)	-0.040 (0.051)	-0.033 (0.049)	-0.106*** (0.012)	-0.092*** (0.007)
Fem*Math Teach Stereotypes		-0.010 (0.007)	-0.007 (0.007)	-0.008 (0.007)	-0.009 (0.008)	-0.007 (0.007)	-0.021** (0.010)	
Math Teach Stereotypes		0.003 (0.005)	0.004 (0.005)	0.002 (0.005)	0.000 (0.006)	0.002 (0.005)	0.016** (0.008)	
Fem*Literature Teach Stereotypes							0.002 (0.010)	-0.003 (0.007)
Literature Teach Stereotypes							-0.001 (0.008)	0.002 (0.006)
Constant	0.283*** (0.005)	0.285*** (0.005)	0.193*** (0.008)	0.121*** (0.040)	0.119*** (0.043)	0.136*** (0.043)	0.289*** (0.008)	0.283*** (0.005)
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R <sup>2</sup>	0.068	0.068	0.109	0.118	0.117	0.229	0.076	0.068
Effect on girls		-0.006 (0.005)	-0.003 (0.005)	-0.007 (0.005)	-0.008 (0.006)	-0.005 (0.006)	-0.005 (0.008)	
School cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	Yes	Yes	No	No

Notes: This table reports OLS estimates of equation 2, where the dependent variable is the high-school track choice. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVI. Estimation of the effect of teachers' gender stereotypes on high-school track recommendation of teachers - school by cohort FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable: Vocational High-School Track Recommendation</b>								
Fem	-0.077*** (0.008)	-0.070*** (0.009)	-0.064*** (0.008)	-0.077 (0.060)	-0.090 (0.062)	-0.111** (0.055)	-0.058*** (0.013)	-0.086*** (0.009)
Fem*Math Teach Stereotypes		0.018** (0.009)	0.013 (0.008)	0.013 (0.008)	0.012 (0.009)	0.004 (0.008)	0.020* (0.011)	
Math Teach Stereotypes		-0.007 (0.006)	-0.008 (0.006)	-0.011** (0.006)	-0.013** (0.006)	-0.014** (0.006)	-0.000 (0.008)	
Fem*Literature Teach Stereotypes							0.003 (0.012)	0.007 (0.009)
Literature Teach Stereotypes							-0.003 (0.009)	-0.011 (0.006)
Constant	0.435*** (0.005)	0.432*** (0.006)	0.516*** (0.010)	0.520*** (0.046)	0.540*** (0.048)	0.382*** (0.045)	0.426*** (0.008)	0.434*** (0.006)
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R <sup>2</sup>	0.093	0.094	0.231	0.234	0.203	0.390	0.092	0.089
Effect on girls		0.011* (0.007)	0.005 (0.006)	0.001 (0.006)	-0.001 (0.007)	-0.009 (0.006)	0.019** (0.009)	
<b>Panel B- Dependent Variable: Scientific High-School Track Recommendation</b>								
Fem	-0.029*** (0.006)	-0.033*** (0.006)	-0.035*** (0.006)	-0.057 (0.045)	-0.060 (0.048)	-0.052 (0.045)	-0.042*** (0.010)	-0.032*** (0.007)
Fem*Math Teach Stereotypes		-0.008 (0.006)	-0.006 (0.006)	-0.009 (0.006)	-0.010 (0.007)	-0.007 (0.007)	-0.015 (0.009)	
Math Teach Stereotypes		-0.001 (0.005)	-0.001 (0.005)	0.000 (0.005)	-0.000 (0.006)	0.000 (0.005)	0.000 (0.007)	
Fem*Literature Teach Stereotypes							0.003 (0.010)	-0.002 (0.007)
Literature Teach Stereotypes							-0.010 (0.008)	-0.006 (0.006)
Constant	0.196*** (0.004)	0.195*** (0.005)	0.121*** (0.006)	0.096*** (0.036)	0.089** (0.040)	0.122*** (0.038)	0.201*** (0.007)	0.204*** (0.005)
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R <sup>2</sup>	0.174	0.174	0.225	0.229	0.229	0.353	0.151	0.158
Effect on girls		-0.010* (0.005)	-0.006 (0.005)	-0.009* (0.005)	-0.010* (0.006)	-0.007 (0.006)	-0.014* (0.008)	
School cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	Yes	Yes	No	No

Notes: This table reports OLS estimates of equation 2, where the dependent variable is the high-school track recommendation of teachers. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVII. Estimation of the effect of teachers' explicit and implicit bias on standardized test score in grade 8 - class FE regression

Dependent Variable	Std Math 8th grade				Std Reading 8th grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.130*** (0.027)	-0.102*** (0.033)	0.028 (0.085)	0.017 (0.086)	0.214*** (0.040)	0.221*** (0.042)	0.397*** (0.097)	0.307*** (0.099)
Fem*Dif InnateAbility(MathTeach)	-0.060** (0.031)	-0.065** (0.027)	-0.067** (0.028)	-0.069** (0.028)				
Fem*Math Teacher Stereotypes				-0.032*** (0.011)				
Fem*Dif InnateAbility(LitTeach)					0.009 (0.042)	0.004 (0.039)	0.016 (0.038)	0.001 (0.038)
Fem*Lit Teacher Stereotypes								-0.001 (0.013)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Teacher Controls	No	No	Yes	Yes	No	No	Yes	Yes
Obs.	30359	30359	30359	30359	29486	29486	29486	29486
R <sup>2</sup>	0.206	0.275	0.276	0.276	0.181	0.291	0.291	0.291

*Notes:* This table reports OLS estimates, where the dependent variable is the standardized test score in grade 8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable “Fem” indicates the gender of the student, “Dif Innate Ability” assumes value 1 if the teacher thinks there are innate differences in math abilities between men and women, 0 if there are few or no differences. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree with honour, refresher courses, age, type of contract, and education of the teacher’s mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVIII. Estimation of the effect of teachers' gender stereotypes on grading by teacher

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A- Dependent Variable</b>	<b>Math Grade end middle school</b>				<b>High Math Grade (&gt;than6)</b>			
Fem	0.391*** (0.015)	0.389*** (0.016)	0.348*** (0.101)	0.348*** (0.101)	0.130*** (0.006)	0.130*** (0.007)	0.097** (0.046)	0.097** (0.046)
Fem* Math Teacher Stereotypes		-0.005 (0.018)	-0.006 (0.018)	-0.006 (0.018)		-0.001 (0.008)	-0.002 (0.008)	-0.002 (0.008)
Constant	6.662*** (0.010)	6.662*** (0.010)	6.563*** (0.024)	6.563*** (0.024)	0.460*** (0.004)	0.460*** (0.004)	0.408*** (0.011)	0.408*** (0.011)
Obs.	21424	21424	21424	21424	21424	21424	21424	21424
R <sup>2</sup>	0.523	0.523	0.532	0.532	0.398	0.398	0.409	0.409
<b>Panel B- Dependent Variable</b>	<b>Literature Grade end middle school</b>				<b>High Literature Grade (&gt;than6)</b>			
Female	0.335*** (0.013)	0.336*** (0.014)	0.314*** (0.027)	0.564*** (0.147)	0.124*** (0.006)	0.125*** (0.007)	0.156*** (0.014)	0.294*** (0.070)
Fem*Lit Teacher Stereotypes		-0.003 (0.013)	-0.003 (0.013)	0.005 (0.015)		-0.003 (0.006)	-0.003 (0.006)	-0.002 (0.007)
Constant	6.752*** (0.008)	6.752*** (0.008)	6.641*** (0.021)	6.640*** (0.021)	0.562*** (0.004)	0.562*** (0.004)	0.502*** (0.011)	0.502*** (0.011)
Obs.	20514	20514	20514	20514	20514	20514	20514	20514
R <sup>2</sup>	0.572	0.572	0.583	0.584	0.412	0.412	0.425	0.425
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sq Std Test 8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes

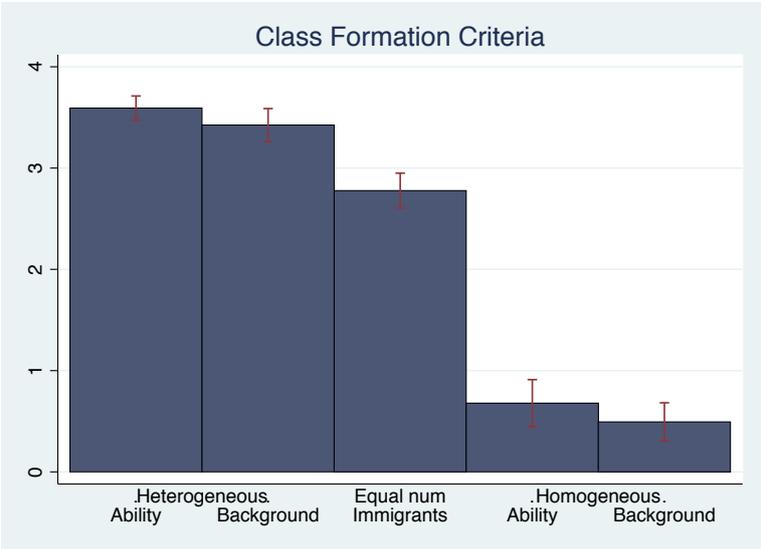
Notes: This table reports OLS estimates of equation 1, where the dependent variable is the grades given by teachers in columns 1-4 and a dummy for a grade higher than 7 in column 4-8. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

## B Survey to School Principals

School principals were asked to complete a paper questionnaire, including information about the career counseling service offered by the school to students, class formation criteria at the beginning of middle school and refresher courses offered to teachers. 82 principals completed the questionnaire. Among them, 94% have a full-time contract, the average experience is 8.2 years, and in 15% of schools the principal is mainly ruling a different complex of school and has been assigned the direction of the institution in our sample temporarily until new staff is hired. This practice is widespread in Italy.

Among 81 principals who completed the questions on class formation, 64 percent consider ability heterogeneity within classes “Extremely important” and 33 percent “Important”. Socioeconomic heterogeneity within classes is considered “Extremely important” by 60 percent of principals and “Important” by 29 percent. The equal allocation of immigrants across classes is considered “Extremely Important” by 25 percent of principals and “Important” by 38 percent. The summary statistics on the importance given to the different criteria in class formation are summarized in Figure B.I.

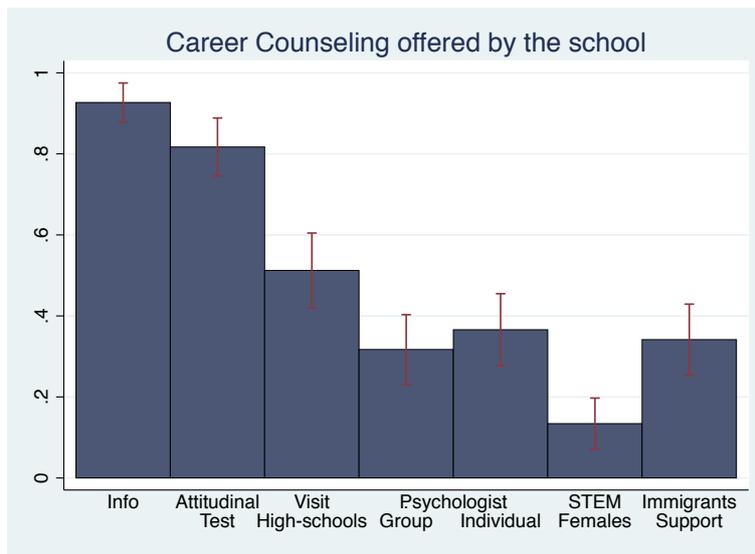
Figure B.I. Class formation criteria according to principals



An entire section of the questionnaire is dedicated to career counseling practices done in the school and the share of schools offering the different services is reported in Figure B.II. Most institutions claim to offer information on high-school curricula and attitudinal tests that help high-school choice. Around one third of schools organize meetings with psychologists at the

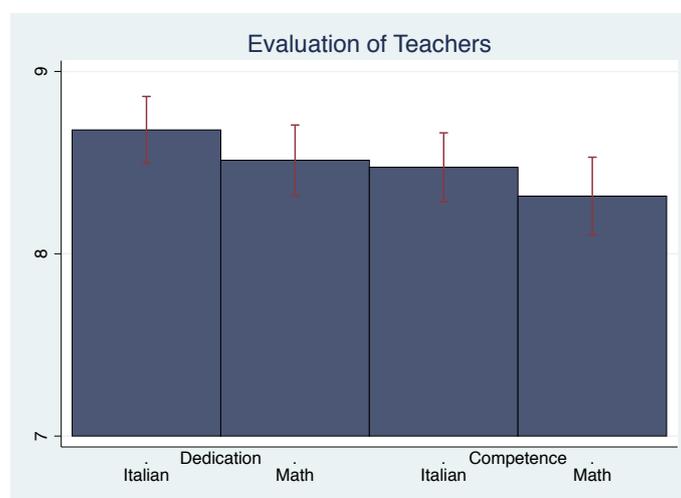
individual or group level to induce students to reflect on this important choice. However, very few institutions try to direct females toward STEM education.

Figure B.II. Share of schools offering different career counseling services



Principals are asked to assign a score from 1 to 10 to the dedication and competence of math and literature teachers in the middle school: generally principals are generous with the evaluation, but, if anything, math teachers are considered marginally less committed to their work and less competent than their literature colleagues, as shown in Figure B.III.

Figure B.III. Dedication and competence of teachers according with principals



## C Teacher Survey

### C.1 Gender Implicit Association Test

The concept behind IAT is that the easier the mental task, the faster the response production and the fewer the errors made in the process.<sup>1</sup>

I invite teachers to complete a seven-block IAT following the schematic overview presented in Figure C.I. Half of the teachers, randomly selected at the individual level completed the IAT as presented in Figure C.I, while the other half completed the IAT with the blocks in the following order: 1, 5, 6, 7, 2, 3, 4 (“order incompatible” IAT). The order of the two tasks is randomly selected at individual level. The blocks used to calculate the IAT score are blocks 3, 4, 6, and 7. The number of words that need to be categorized is 20 in blocks 3 and 6, and 40 in blocks 4 and 7, as in the standard IAT 7-blocks. The measure of implicit stereotypes is calculated as the difference in reaction time between the task in which scientific fields and male names are on the same side of the screen and the task in which scientific fields and female names are on the same side of the screen. The scoring procedure follows the guidelines of the improved scoring algorithm defined by Greenwald et al. (2003).

Blocks	Left Categories	Right Categories
1	Maschio (Male)	Femmina (Female)
2	Scientifico (Scientific)	Humanistic (Umanistico)
3	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
4	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
5	Humanistic (Umanistico)	Scientifico (Scientific)
6	Maschio (Male) Humanistic (Umanistico)	Femmina (Female) Scientifico (Scientific)
7	Maschio (Male) Humanistic (Umanistico)	Femmina (Female) Scientifico (Scientific)

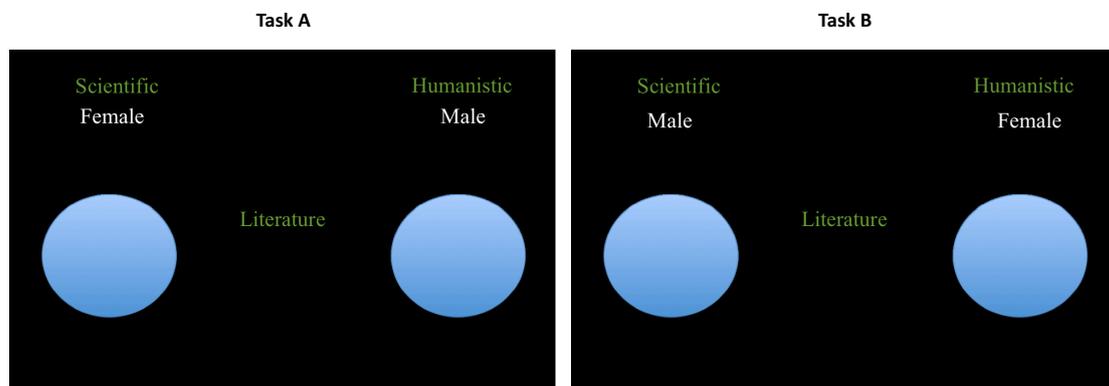
Figure C.I. Schematic overview of the Gender Implicit Association Test

Stimuli presented within each category are summarized in Figure C.II, while a screenshot of the tablet is shown in Figure C.III.

<sup>1</sup>It was initially developed by Donders (1868). Donders was very optimistic about the possibility of quantifying how the mind works using the “time required for simple mental processes” and performed some of the first experiments making participants respond with the right hand to stimuli on the right side and with the left hand to stimuli on the left side.

Categories	Stimuli
Maschio (Male)	Luca, Federico, Matteo, Alberto, Davide, Alessandro
Femmina (Female)	Anna, Martina, Laura, Giulia, Chiara, Alessia
Scientifico (Scientific)	Matematica (Math), Fisica (Physics), Scienze (Science), Chimica (Chemistry), Ingegneria (Engineering), Calcolo (Calculus)
Humanistic (Umanistico)	Lettere (Literature), Italiano (Italian), Filosofia (Philosophy), Letteratura (Literature), Storia (History), Lingue (Languages)

Figure C.II. Category Labels and Stimuli for the Implicit Association Tests



IAT score=0	No Bias	Time/Errors of Task A = Time/Errors of Task B
IAT score>0	Bias Pro Boys	Time/Errors of Task A > Time/Errors of Task B
IAT score<0	Bias Pro Girls	Time/Errors of Task A < Time/Errors of Task B

Figure C.III. Screenshot of the Implicit Association Test for the two types of tasks

Teachers were asked to complete a race IAT with male names and female names of Italians and immigrants. The order of race and gender IATs was randomized at the individual level. In Table C.I, I check the influence of order of blocks on the IAT score. On average, there is a small difference in the IAT score between individuals who perform the order compatible and incompatible test. Hence, in all regressions where there are no class (and therefore teacher) fixed effects, I control for the order of IATs.

Table C.I. Correlation between Gender-Science IAT score and order of different parts of the survey

Dependent variable : IAT score Math Teachers					
	(1)	(2)	(3)	(4)	(5)
First IAT Gender	0.048** (0.023)			0.049** (0.023)	0.050** (0.025)
First Questionnaire, then IAT		-0.093** (0.045)		-0.096** (0.046)	-0.078 (0.055)
Order Compatible IAT Gender			-0.066*** (0.021)	-0.069*** (0.021)	-0.069*** (0.022)
School FE	No	No	No	No	Yes
Obs.	1378	1378	1378	1378	1378
$R^2$	0.108	0.106	0.111	0.116	0.170

*Notes:* This table reports OLS estimates of the correlation between order of IAT and IAT score. A higher value of IAT score means stronger implicit association between Male-Science and Female-Literature. The dummy “First IAT Gender” captures the order of IATs (gender and race). The variable “Order Compatible IAT Gender” captures whether a teacher was asked first to associate compatible categories (Male-Scientific vs. Female-Humanistic) or the opposite (Female-Scientific vs. Male-Humanistic). Finally, in 8 cases for math teachers and 39 cases for literature teachers, I asked them to first complete a questionnaire and then the IATs. All columns include a dummy variable which assumes value 1 for math teachers. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

## **C.2 Teachers' Questionnaire**

### **Factors influencing track choice**

*“Female students with the same math grade of males are less likely to attend a scientific track during high-school. According with your experience, how much can these factors influence the choice of females toward alternative tracks?”* Answer in a scale from 1 to 5, where 1 means ‘*Not at all*’ and 5 means ‘*A lot*’.

1. Low interest for scientific subjects
2. Low inclination for scientific subjects
3. Low self-esteem
4. Encouragement of the family toward alternative paths
5. Influence of gender predicament (“women are bad at math”)

### **Factors influencing grading**

*“When you grade your students, which weight do you assign to the following components?”* Answer in a scale from 1 to 5, where 1 means ‘*A little*’ and 5 means ‘*A lot*’.

1. Performance in written exams
2. Performance in oral exams
3. Diligence in doing homeworks

### **Factors influencing track recommendation**

*“When you give the high-school track recommendation to your students, which weight do you assign to the following components?”* Answer in a scale from 1 to 5, where 1 means ‘*A little*’ and 5 means ‘*A lot*’.

1. Grades and performance at school
2. Predisposition and interests of the student
3. Parents' education
4. Economic resources of the family
5. Engagement of family in schooling

### **Teachers' Track recommendation**

*“When you give the high-school track recommendation, how much weight do you assign to these aspects?” Answer in a scale from 1 to 5, where 1 means ‘Not at all’ and 5 means ‘A lot’.*

- Performance at school
- Attitude and interest of students
- Education of parents
- Economic resources of families
- Participation and involvement of parents
- Presence of specific programs for immigrants (e.g. Italian courses)

### **Explicit gender bias**

*Do you agree with the following sentences?*

- *There are innate biological differences in math abilities of women and men: ‘Not at all’, ‘A little’, ‘A bit’, ‘A lot’, ‘Absolutely’*
- *When jobs are scarce, men should have more right to a job than women: ‘Agree’, ‘Neither Agree nor Disagree’, ‘Disagree’*

### **C.3 Students' Questionnaire**

*“Evaluate for each school subject your ability to tackle it.”: ‘Good’, ‘Mediocre’, ‘Scarce’*

## D Effort vs. Talent

Research in social psychology shows that school teachers with higher gender stereotypes believe that males are more talented than females, even among students with comparable level of ability (Tiedemann, 2002). The performance of females is mainly attributed to effort. Furthermore, teachers are more likely to convey their stereotyping of *mathematics as a male domain* through their classroom instruction (Keller, 2001; Riegle-Crumb and Humphries, 2012) compared to *literature as a female domain*.

In this Appendix, I provide evidence that teacher perception of gender differences in effort and ability mirrors (or is mirrored by) pupils own perception. I have information on a sample of around 17000 Italian students in grade 6 collected by the Italian National Evaluation Center (INVALSI) on the reasons why students believe they are performing well in both math and reading. In particular, students are asked: “You correctly solved all exercises in a math exam. Why does it happen?” and “You present to your classmates a text you have read. You do it in a clear and precise way and everybody follows it with interest. Why does it happen?” Students can choose among different potential answers: “I was helped”, “I was lucky”, “It was easy”, “I am good” and “I put in a lot of effort”. Here I focus on the latter two answers. Female students are 7.6 and 2.3 percentage points less likely to say “I am good” in math and reading respectively than males (column 1 and 2, Table D.1). The difference in under- confidence between the two subjects is statistically significant for females, even controlling for individual level fixed effects (column 3). Finally, in column 4, I show that females are more likely than males to believe the most important reason of their success is effort. However, the difference between subjects is slightly lower, although statistically significant at 1 percent level. Unfortunately, in this dataset we do not have information on teacher gender beliefs to access whether they affect the pupil own beliefs about the reason of successful performance in math and reading.

Table D.I. Correlation between subject, gender, and reasons for performing well in exams

Dependent Variable: Reasons for performing well						
	Being good			Exerting effort		
	Math	Reading	Reading and Math	Math	Reading	Reading and Math
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.076*** (0.005)	-0.023*** (0.006)		0.139*** (0.007)	0.089*** (0.007)	
Test math	0.031*** (0.002)			0.010*** (0.002)		
Test reading		0.005*** (0.002)			0.049*** (0.002)	
Math			-0.026*** (0.009)			0.080*** (0.010)
Math*Female			-0.062*** (0.011)			0.031** (0.013)
Individual FE	No	No	Yes	No	No	Yes
Obs.	16952	16979	34028	16952	16979	34028
R <sup>2</sup>	0.042	0.001	0.531	0.025	0.038	0.605

Notes: The specific questions asked at the end of the standardized test in grade 6 for the cohort of students graduating from middle school in 2013-14 are: (1) *In a math exam, you correctly solved all exercises. Why does it happen?* (2) *You present to your classmates a text you have read. You do it in a clear and precise way and everybody follows it with interest. Why does it happen?* The potential answers are: a) *I was helped* b) *I was lucky* c) *It was easy* d) *I'm good* e) *I put in a lot of effort*. Robust Standard Errors clustered at class level are in parentheses. This information is collected together with the standardized test score in grade 6. "Female" refers to the gender of students. "Math" is a dummy which assumes value 1 if the dependent variable is related to own assessment in math and 0 if it is related to own assessment in reading. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

## E Conceptual Framework

I develop a simple conceptual framework, similar to [Dee \(2014\)](#) and based on the stereotype threat idea developed in social psychology that can help interpreting the results. In this model, student beliefs about own ability in math are a function of own true unobserved ability  $a_i$  and teachers' gender stereotypes  $s_t$ , defined as follows:  $\alpha_i = f^i(a_i, s_t)$ . The impact of the bias on own self-perception is an individual specific function: students with higher vulnerability to the stereotype that "girls are not good at math" will be more negatively impacted by teacher stereotypes. This simple framework is flexible enough to capture heterogeneous perception of own ability among students with the same true unobserved ability  $a_i$ . For instance, a boy may perceive higher ability in math compared to a girl with the same unobserved talent for math  $a_i$  in the same class (i.e., exposed to the same teacher with stereotypes  $s_t$ ). I assume that students' beliefs about own ability are a weakly decreasing function of teacher stereotypes, i.e.,  $\alpha_s \leq 0$ . This is a testable assumption, which is supported by the empirical evidence available in [Section 6](#) of the paper.

In a simple framework, students choose effort and individual's utility can be represented as

$$u_i = \theta_i k(\alpha_i, e_i) - c(e_i) \quad (1)$$

where  $u$  is differentiable and the sufficient conditions for a local interior maximum hold,  $k$  is the benefits function, which depends on ability ( $\alpha_i$ ) and effort ( $e_i$ ), and the cost  $c$  is paid according with the level of effort  $e$  exerted by individual  $i$ . The component  $\theta_i$  introduces an exogenous heterogeneity and it captures observable difference across individuals in the returns to performance. In this simple framework, I do not introduce parametric assumptions on the utility function.

I am interested in how the optimal level of effort of students varies with stereotype of the teacher. The model implies that:

$$e_s^* = \frac{\theta k_{e\alpha} \alpha_s}{-(\theta k_{ee} - c_{ee})} \begin{matrix} \geq 0 \\ < 0 \end{matrix} \quad (2)$$

The second order condition for a relative maximum implies that the second order derivative must be negative and therefore the denominator in equation (2) must be positive. Furthermore, I assumed that  $\alpha_s \leq 0$ , which implies that higher teacher stereotypes have a negative or null impact on self-perception of student  $i$ 's ability, *ceteris paribus*. Hence, the optimal level of effort with respect to stereotype ( $e_s^*$ ) depends on the complementarity or substitutability of effort and perceived ability ( $k_{e\alpha}$ ).

Effort and perceived ability are often considered as complementary in the education production function (i.e.  $k_{e\alpha} > 0$ ), so that a higher self-assessment of own capacities enhances the motivation to exert effort (Bénabou and Tirole, 2002). Hence, higher stereotypes will decrease the level of effort in equilibrium ( $e_s^* \leq 0$ ). However, if the student increases effort as a reaction to a negative stance of the math teacher (i.e.  $k_{e\alpha} < 0$ ), then the impact of stereotypes on effort is positive. As suggested also by Dee (2014),  $e_s^* \geq 0$  is likely, for instance if individuals of the stigmatized group consider the stereotype strongly improper and react with an “*I’ll show you are wrong*” attitude. In the context of gender stereotypes, it would imply that talented female students may increase the level of effort when they interact with teachers with stronger bias in order to disprove the negative belief.

In the empirical counterpart of this model, I observe improvements in achievement test scores ( $P$ ) and not directly effort<sup>2</sup>, but I assume for simplicity that the derivative of performance with respect to effort is positive ( $P_e > 0$ ) and I focus on the choice of the latter. Indeed, in the paper, I analyze whether improvements in achievement test scores are affected by teacher stereotypes. Assume two students, with the same gender, family background and math performance, are quasi-randomly assigned to two different teachers, respectively with stereotypes  $s_{t_i}$  and  $s_{t_j}$ , such that  $s_{t_i} < s_{t_j}$ . Then, if effort is complementary of students’ perceived ability, the optimal level of effort (and therefore performance) of the student decreases with teacher stereotypes. However, if  $k_{e\alpha} < 0$ , then  $e_s^* > 0$ . This theoretical result may explain why girls at the top of the initial ability distribution have slightly higher, albeit indistinguishable from zero, improvements in math when exposed to teachers with stronger gender stereotypes.

## E.1 Extension of the Conceptual Framework

I extend the simple conceptual framework to analyze the impact of teachers’ gender stereotypes on effort of students, as mediated by both student perception of own ability ( $\alpha_i$ ) and teacher investment toward pupils, in the form of either time or encouragement, ( $\beta_{t_i}$ ). The latter is an additional channel through which teacher bias ( $s_t$ ) may impact students’ performance and choices. I define  $\beta_{t_i}$  as an individual specific function of ( $s_t$ ):  $\beta_{t_i} = h^i(s_t)$ . Furthermore, I assume that teachers with higher stereotypes are less supportive toward member of the stigmatized group, i.e.,  $\beta_s \leq 0$ . Unfortunately, I do not observe data on gender specific investment or interaction in the classroom between professors and students, but the social psychology literature described in Section 6 provides evidence in support of this assumption.

---

<sup>2</sup>Even ideally having information about the number of hours studied, it is not clear that this is necessarily a better measure of effort since the quality of time use is also essential in the learning process.

The individual chooses the level of effort in order to maximize:

$$u_i = \theta_i k(\alpha_i, e_i, \beta_{t_i}) - c(e_i) \quad (3)$$

where  $\beta_{t_i} = h^i(s_t)$  and all other parameters and functions are defined as in equation (2).

The optimal level of effort with respect to teacher stereotypes is given by:

$$e_s^* = \frac{\theta(k_{e\alpha}\alpha_s + k_{e\beta}\beta_s)}{-(\theta k_{ee} - c_{ee})} \begin{matrix} \geq 0 \\ \leq 0 \end{matrix} \quad (4)$$

In this extended framework, whether students increase or decrease their effort level when exposed to more biased teachers will depend on the complementarity and substitutability of effort with both own perceived ability ( $k_{e\alpha}$ ) and teacher behaviour  $k_{e\beta}$ . If both are complementary, then the student will decrease the level of effort when exposed to a teacher with higher bias ( $e_s^* < 0$ ). If both are substitutes, we are in the case in which students work harder when exposed to more biased teachers to disprove the negative belief.

## **F Additional Outcome: Retention rate**

In the Italian schooling system, at the end of each academic year, teachers decide whether the student is admitted to the following grade. This decision is based on the overall assessment of students, including both performance and behavior in class. The retention rate of males is higher compared to the one of females. For instance, among students in my sample, 7.7% of males and 4.7% of females are retained in at least one of the three years of middle school. In Table [F.I](#), I check whether math teachers' bias has an impact on retention rate in the last year of middle school, when the probability of school failure is on average 0.2% for boys and 0.1% for girls. I find suggestive evidence that girls in classes assigned to math teachers with more gender stereotypes are slightly more likely to fail (Panel A), while there is no significant effect for literature teachers (Panel B). However, these effects are driven by very few cases.

Table F.I. Estimation of the effect of teachers' gender stereotypes on grade retention in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)
<b>Panel A- Dependent Variable: grade retention at the end of grade 8</b>					
Female	-0.001* (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.004** (0.002)	-0.007* (0.004)
Fem*Math Teacher Stereotypes		0.001* (0.000)	0.001* (0.000)	0.001* (0.000)	0.001* (0.000)
Constant	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
Obs.	24840	24840	24840	24840	24840
R <sup>2</sup>	0.076	0.076	0.077	0.077	0.078
<b>Panel B- Dependent Variable: grade retention at the end of grade 8</b>					
Female	-0.001** (0.001)	-0.001* (0.001)	-0.001* (0.001)	-0.005*** (0.002)	0.000 (0.004)
Fem*Lit Teacher Stereotypes		-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.001)
Constant	0.002*** (0.000)	0.002*** (0.000)	0.003*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Obs.	23804	23804	23804	23804	23804
R <sup>2</sup>	0.073	0.073	0.074	0.075	0.075
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

*Notes:* This table reports OLS estimates of equation 1, where the dependent variable is math or reading standardized test score in grade 8 in Panel A and B, respectively, and the dependent variables refer to math teachers in Panel A and to literature teachers in Panel B. The unit of observation is student  $i$  in class  $c$  taught by teacher  $t$  in grade 8 of school  $s$ . Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. \*, \*\*, and \*\*\* indicate significance at the 10%, 5% and 1% percent level respectively.

## References

- Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics* 117(3), 871–915.
- Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry* 52(1), 173–182.
- Donders, F. (1868). *On the speed of mental processes. Translation by WG Koster in Attention and performance II, ed. WG Koster.* North Holland.
- Greenwald, A. G., B. A. Nosek, and M. R. Banaji (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology* 85(2), 197.
- Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1), 5–86.
- Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology* 141(2), 165–173.
- Riegle-Crumb, C. and M. Humphries (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society* 26(2), 290–322.
- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics* 50(1), 49–62.