Implicit Stereotypes: Evidence from Teachers' Gender Bias

Faculty Research Working Paper Series

Michela Carlana
Harvard Kennedy School

October 2018

RWP18-034

Implicit Stereotypes: Evidence from Teachers' Gender Bias*

Michela Carlana†

July 12, 2018

Abstract

I study whether exposure to teachers' stereotypes, as measured by the *Gender-Science Implicit Association Test*, affects student achievement. I provide evidence that the gender gap in math performance substantially increases when students are assigned to teachers with stronger gender stereotypes. Teachers' stereotypes induce girls to underperform 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. The findings are consistent with the hypothesis that stereotypes impair the test performance of ability-stigmatized groups, who end up failing to achieve their full potential.

JEL: J16, J24, I24.
Keywords: gender, math, teachers, implicit stereotypes, IAT, self-confidence, track choice.

*I thank for their useful suggestions Alberto Alesina, Ingvild Almas, Thomas Le Barbanchon, Pamela Giustinelli, Selim Gulesci, Eliana La Ferrara, Giampaolo Lecce, Nicola Gennaioli, Jennifer Lerner, Andreas Madestam, Valerio Nispi Landi, Laura Ogliari, Paolo Pinotti, Jonathan de Quidt, Dan-Olof Rooth, David Stromberg, Guido Tabellini, Marco Tabellini, Anna Tompsett, Diego Ubfal and seminar participants of 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. 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 14th September 2016.

†Harvard Kennedy School (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, up to a reversal of the gap in school attainment in many contexts (Goldin et al., 2006). In spite of this, boys outperform girls in math in most countries and the gender gap in favour of boys is even wider among the highest-achieving students (OECD, 2014). A long-standing debate attributes gender differences in mathematics to either biologically based explanations in brain functioning or culture and social conditioning (Baron-Cohen, 2003; Ceci et al., 2009; Nollenberger et al., 2016). Cross-countries evidence supports the latter idea by showing that the more gender equality a country has, the smaller its gender gap in math (Guiso et al., 2008; Nosek et al., 2009; Else-Quest et al., 2010).[1] Gaining a better understanding of whether stereotypes affect the emergence of the gap in math skills is important to potentially explain the enduring gender differences in readiness for science, technology, engineering, and math (STEM) universities and the underrepresentation of women in highly profitable occupations and fields (Altonji and Blank, 1999; Card and Payne, 2017).

The gender stereotypical belief that women are worse than men in scientific fields is pervasive and deeply-held in most societies and by most individuals, including parents and teachers (Tiedemann, 2000; Reuben et al., 2014; Bordalo et al., 2016). In this paper, I study whether exposure to gender stereotypes of teachers, as measured by the *Gender-Science Implicit Association Test*, affects educational achievement of boys and girls. Using unique Italian data including detailed first-hand surveys and administrative information, I find that teachers with stronger gender stereotypes have a negative and quantitatively significant influence on girls. First, I show that the gender gap in math improvements during middle school increases by 38 percent when students are assigned to teachers with one standard deviation higher implicit stereotypes. Said differently, the gender gap in math improvements triples in classes assigned to a math teacher that implicitly associates more boys than girls with mathematics compared to classes assigned to a teacher that has the opposite implicit associations. Teachers' stereotypes have no effect on boys', while they lower girls' math performance, especially of those from disadvantaged background. Second, I provide evidence that teachers' stereotypes induce girls to self-select into less demanding tracks, following the biased recommendation of their teachers. Finally, I show that teachers' stereotypes have a substantial negative impact on girls' self-confidence in math. The findings are consistent with the hypothesis that stereotypes impair the test performance of ability-stigmatized groups, who end up failing to achieve their full potential. Implicit stereotypes create a self-fulfilling prophecy, perpetuating gender differences in

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

math performance.

One of the main challenges to study the impact of teachers' gender stereotypes on educational outcomes is the availability of an appropriate measure of stereotypes, matched with students' achievement and choices. I measure stereotypes by administering the Gender-Science Implicit Association Test (IAT) to around 1,400 math and literature teachers, working in 103 schools in the North of Italy. 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 (Lowes et al., 2015; Burns et al., 2016; Glover et al., 2017). The test exploits the reaction time to associations among 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). Although there is some mixed evidence on its predictive validity (Blanton et al., 2009; Oswald et al., 2013), implicit bias has been found to correlate with some outcomes in the real world and in laboratory experiments, related for instance to hiring decisions (Rooth, 2010; Reuben et al., 2014). In addition to IAT scores, I collected detailed information on teacher characteristics, such as family background, teaching experience and explicit gender beliefs. To perform the analysis, I build a unique dataset, combining these surveys to teachers 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 a newly collected student questionnaire. Data on pupils include performance in math and reading standardized test scores, family background, high-school track choice, teachers' track recommendation and – for a sub-sample of students – also 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 level of implicit stereotypes. I provide supporting evidence showing that baseline characteristics of students, such as family background and initial standardized test scores, are not systematically correlated with teachers' stereotypes. First, I investigate the impact of teachers' stereotypes on the gender gap within the class. I add class fixed effects, which absorb all characteristics of peers, school environment, and teachers, including the level of gender stereotypes. I exploit variations in performance and track choice between boys and girls enrolled in the same class.[2] Second, I compare students of the same gender, enrolled in the same school and cohort, but assigned to teachers with different level of stereotypes. This exercise permits to understand whether the wider gender gap in classes assigned to teacher with more stereotypes is due to girls lagging behind, boys improving more, or a combination of the two

[2]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 exception due, for instance, to retirement or transfer to a different school.

effects.

This paper makes three contributions. First, I collect a dataset including IAT scores and detailed additional information on around 1,400 teachers and I show that implicit associations correlates with observable characteristics, such as gender, field of study, and gender norms in the place of birth, as measured by the World Value Survey and by female labor force participation at province level. This correlation supports the view that IAT scores reflect exposure to cultural stereotypes (Arkes and Tetlock, 2004). Furthermore, I find that IAT scores do not correlate with variables such as gender of own children, teacher quality, experience, and with self-reported gender bias, either because they relate to two different mental constructs or because there is social desirability bias in the explicit answers (Greenwald et al., 2009). Second, the paper provides evidence on the relevance of cultural and social conditioning in affecting the gender gap in math achievement and high-school track choice. More precisely, it uncovers the role of implicit stereotypes in the context of education economics and pupil-teacher interactions that was not previously investigated. This result provides a link between teachers' stereotypes and teacher bias[3]: it suggests 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 the high-school track to their students. Third, I show the influence of teachers on self-confidence on own math ability. This is a crucial channel to explain the underperformance of girls in math when assigned to more biased teachers, but also relevant per se since it suggests that the lower self-confidence of women in the scientific fields is at least partially activated by exposure to gender stereotypes.

This study adds to the recent literature in economics that has underlined the benefits from interacting with social psychologists and considering *implicit bias* in studying discrimination (Guryan and Charles, 2013; Bertrand and Duflo, 2017). Implicit stereotypes can operate even without awareness or intention to harm the stigmatized-group (Nosek et al., 2002; 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; Greenwald et al., 2009; Oswald et al., 2013). I provide evidence that teachers with stronger implicit stereotypes negatively affect math achievements of their female students. They are biased against girls in their track recommendation, with

[3]Stereotypes are overgeneralized and simplified representation of differences between groups, which may hold a *kernel-of-truth* (Bordalo et al., 2017). Indeed, the belief that *women are worse than men in math* is based on a true empirical fact: girls lag behind in math test-scores in most countries by the age of 14. On the other hand, an individual is biased only if his or her own preconceived idea interferes with the ability of being impartial and objective.

long-run implications for their educational career.

Teachers matter for students' performance and their later-life outcomes (Chetty et al., 2014a,b) – and their gender stereotypes may be an important channel. The economics literature analyzing the impact of gender stereotypes of teachers 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, 2015; Terrier, 2015; Lavy and Megalokonomou, 2017).[4] Compared to other measures of teacher bias, the Implicit Association Test has two main advantages. First, it does not suffer from social desirability bias that may be an issue in self-reported measures. Second, teachers' stereotypes are measured without relying on data on student performance, which may capture variation in unobservable characteristics of boys and girls, potentially correlated with future outcomes of pupils. Finally, a growing number of papers exploits the gender of teachers as proxy of exposure to stereotypes of their pupils 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 Gender-Science IAT scores, but the impact of implicit stereotypes on student outcomes is similar in terms of magnitude for male and female teachers.

Finally, I contribute to understanding the importance of gender-biased environments in explaining the under-confidence of females in STEM fields. Gender differences in confidence and competitiveness have negative consequences for women's performance, educational and occupational choices (Coffman, 2014; Reuben et al., 2015; Kugler et al., 2017). Exposure to biased teachers activates negative self-stereotypes on female students. The results are consistent with the predictions of the stereotype threat theory (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).

This paper is organized as follows. Section 2 explains the setting and provides information on the Italian institutional background. Section 3 describes the data available on both students and teachers. Section 4 presents the estimation strategy and tests the identifying assumption of "as good as random" assignment of girls and boys to teachers with higher or lower level of stereotypes. The main results of the paper are reported in Section 5: girls underperform in math and are induced to undertake less demanding high-school tracks when they are exposed to biased teachers. Mechanisms are discussed in Section 6 with a focus on the role of self-confidence. Finally, Section 7 concludes. All supplementary material is provided in the Online Appendices.

[4] Lavy and Megalokonomou (2017), using a panel dataset, show that gender bias in grading of teachers is persistent over time and it influences students' university choice.

2 Setting

In the Italian educational system, middle schools lasts three years, from age 12 to 14. Students in middle school are assigned to classes at the beginning of grade 6 and they stay with the same peers until the end of grade 8.[5] 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.[6] The general criteria mentioned by most schools are equal allocation of students across classes according to gender, disability, socio-economic status, and ability level (as reported by the elementary school). Moreover, I collect information directly from principals on how classes are formed. School principals report that the most relevant aspect in the class formation process is the comparability across classes and heterogeneity within class in the same school (for detailed information, see Online Appendix B). What is important for my analysis is that I can also 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 determined by experience in a centralized system. Teachers' allocation across school is settled by seniority: when they accumulate years of experience, they tend to move close to their home town and away from disadvantaged areas (Barbieri et al., 2011). Each class is assigned by the principal to a math and Italian teacher among those available in the school and they 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 Italian teacher.[7]

Standardized test score in math and reading are administered in grade 2, 5, 6, 8 and 10 by the National Institute for the Evaluation of the Italian Education System (INVALSI).[8] 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 for the one in grade 8. The achievement test score of grade 8 is the highest stakes among these test scores, since until 2017 it affected 1/6 of the final score

[5]There are only few exceptions: students may be transferred to a different school by their parents or be required by their teachers to repeat a grade (overall less than 6% of students).

[6]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, while all schools in my sample are from the North.

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

[8]The test score in grade 6 was administered only up to the school year 2012-13.

of students at the end of middle school. However, this final grade has no direct impact on the enrollment in high-school or on the future educational career of students.

After middle school, students self-select into three different tracks: academic oriented ("liceo"), technical, and vocational high-school. Each type of school is divided in 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, while the vocational track can have different core subjects, for instance hospitality training, cosmetics and mechanical workshop. Students are free to choose a high-school with no restriction on the track based on grades or ability and they tend to choose according to family background and child's enjoyment of the curriculum (Giustinelli, 2016). Teachers give a non-binding track recommendation to families with an official letter sent to children's home, which is also reported to the Ministry of Education.

The choice of high-school is strongly correlated with the 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 vocational track, only 1.7% of the cohort graduating in 2016 enrolled in university, while the percentage increases to 73.7% and 32.3% in the academic and technical track respectively. Interestingly, among students of the technical track the majority enrolls 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.[9] The sample was designed including all schools of the provinces of Milan, Brescia, Padua, Genoa and Turin with more than 20 immigrants in the school year 2011-12 enrolled in grade 6.[10] 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 cohorts. Although the standardized difference is always below 0.25 (Imbens and Rubin, 2015),

[9]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). One school has only one class and missing data for the standardized test score. The final sample used in the analysis includes 90 schools.

[10]The data collection was conducted also for an ongoing work in which we study teacher race bias (Alesina et al., 2018).

as expected, the sample used in this paper has a higher share of immigrants compared to the national average (20.2% vs. 10.0%) and also compared to the average of the five provinces in the North (20.2% vs. 14.3%).[11] Interestingly, the average score in standardized test in mathematics of both boys and girls are similar to the local and national average.

3.2 Data Sources

I use four sources of data: teacher survey data, student survey data, administrative information from the Italian Ministry of Education (MIUR) and from the National Center for the Evaluation of the Italian Educational System (INVALSI). I collected detailed information on teachers, including implicit stereotypes measured by the Gender-Science Implicit Association Test (IAT), and on students' self-assessment of own ability in different subjects.

A1. Implicit Association Test

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 choice. 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.[12] The survey is divided into two parts: the Implicit Association Test (IAT) and a questionnaire.

I measure implicit gender stereotypes using the IAT, a tool developed by social psychology (Greenwald et al., 1998; Lane et al., 2007). The idea underlying the test is that the easier the mental task, the faster the response production and the fewer the errors made in the process.[13] The IAT requires the categorization of words to the left or to the right of a computer or tablet screen and it provides a measurement of the strength of the association between two concepts –

[11] Around half the students are first generation and half are second generation immigrants.

[12] 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 of schools with an attendance rate of teachers to the survey higher than 70% with summary statistics on the outcomes of their students during high-school and they were strongly interested in it.

[13] This concept was initially developed by Donders (1868). Donders was very optimistic about the possibility of quantifying how 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.

specifically in the Gender-Science IAT, 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). One word at a time appears at the center of the screen and individuals are instructed to categorize them as fast as possible to the left or the right according with 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 to one side of the screen male names and scientific subjects and to the opposite side of the screen 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 order of the two tasks is randomly selected at individual level. 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. The measure of implicit stereotypes is calculated as the difference in reaction time in the task in which scientific fields and male names are in the same side of the screen compared to the task in which scientific fields and female names are in the same side of the screen. Detailed explanation of the IAT is provided in the Online Appendix C.[14]

A broad strand of literature in social psychology and an increasing number of papers in economics have 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 in a lab experiment that higher stereotypes (measured by the Gender-Science IAT) predict employers' biased expectations against female math performance and also 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 self, even for women who had selected math-intensive majors (Cvencek et al., 2011; Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007). 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).

There is a lively debate among social psychologists on Implicit Association Tests. First, some papers have argued that IAT has weak predictive validity (Blanton et al., 2009; Oswald

[14]The order of the tasks was randomized at individual level and in the Online Table A.II 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.

et al., 2013). Most of the studies refer to experiment with less than 50 subjects and they don't have information outside the lab on whether individuals with stronger implicit associations are actually biased in their interaction with stigmatized groups. Hence, I believe that further research is necessary and my paper can contribute to this debate. Second, some studies suggest that IAT can be faked after respondents acquired knowledge of the test (Fiedler and Bluemke, 2005). IAT is not widespread in Italy and none of the teachers who took the survey reported to be familiar with the test. Without any hints, it seems unlikely that they were able to figure out how to trick the test. However, if they were, this would increase the noise in the IAT score and lead to an attenuation bias when estimating the impact of implicit stereotypes on student outcomes. Third, IAT scores could be contaminated by *extrapersonal associations* that are are available in memory, but that do not contribute to an individual's personal evaluation when one interacts with the specific category (Olson and Fazio, 2004). I design a IAT test specifically related to the schooling context: teachers complete the survey within the school building by associating school subjects and gender. The concern of capturing associations outside the schooling context is therefore alleviated. Forth, at least part of IAT scores are capturing unstable characteristics that do vary over time. 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). This short-term exposure may introduce additional noise in the measurement. Finally, some papers in social psychology argue that "the data may reflect shared cultural stereotypes rather than personal animus" (Arkes and Tetlock, 2004). This is coherent with my finding of 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 relevance of implicit stereotypes.

To sum up, IAT scores are noisy measure of implicit stereotypes that may be affected by culture and socialization experience. Nevertheless, they have the great advantage of avoiding social desirability bias in the response and capturing implicit association 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 teacher have stereotypes (i.e. in the level of IAT score), but on whether those with higher stereotypes have a negative impact on performance, track choice, and self-confidence of girls and boys.

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, etc). Furthermore, they were also asked questions about explicit bias, as 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"*.[15] 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 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 in their websites.

B. Administrative Data and Students' Self-Confidence

I obtained individual level information from the Italian Ministry of Education and from the National Institute for the Evaluation of the Italian Education System (INVALSI) for three cohort of students enrolled in grade 6 between school year 2010-11 and 2012-13.[16] The data available include math and reading standardized test score in grade 6 and 8, parents' education and occupation, baseline individual information (date and place of birth, gender, citizenship), high-school track choice and official teachers' recommendation. Students in grade 8 in 2014 of 24 schools in this sample are asked to complete a survey about their track choice, around two months before the end of middle schools. In particular, they need to report their belief about their own ability in each subject, choosing between "good", "mediocre", "scarce".[17]

3.3 Descriptive Statistics

Math Teachers

The dataset includes 537 math teachers, but I have to restrict the main analysis to 303 teachers ("matched sample") who were working for the same school even before 2016 and for which I have student data.[18] Online Appendix Table A.III shows the balance table of the differences between the sample of teachers matched (303 teachers) and the other 234 math teachers who completed the IAT. As expected, teachers not matched are around 9 years younger, 40 percent

[15] The specific questions are reported in the Online Appendix C.2.

[16] Individual level data are anonymous and I obtained the authorization from each school principal to access data from their school.

[17] The specific question is reported in the Online Appendix C.2.

[18] I have information also on 853 Italian teachers, but the main focus of this paper is on math teachers given the strong belief that men are better than women in mathematics. Part of the results exploiting data on literature teachers are available in the Online Appendix. Further information available upon request to the author.

less likely to have full-time contract, and they have 12 years less 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: 0.971, Online Appendix Figure A.I).

Table I reports descriptive statistics. Most teachers are females (84%), they are on average 52 years old with 23 years of experience in teaching and 92% hold a full-time contract. The majority (66%) of math teachers are born in a city in the North of Italy, but a substantial share is born in the Center or South of Italy and then migrated to the North to work. Most teachers graduated from programs in biology, natural sciences and other related subjects: only 24% studied math, physics, or engineering. At the bottom of Table I, I report the summary statistics of explicit bias questions described in details 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.

Considering the IAT thresholds typically used in the social psychological literature, 24% of teachers slightly or moderately associates math with girls, 31% presents little to no clear associations, 19% shows slight male-math association and 26% show moderate to severe male-math associations.[19] For comparison, the sample of 1164 Italians used by Nosek et al. (2009) have an average Gender-Science IAT score of 0.40 (SD 0.40): the score of math teachers is on average lower (mean 0.09, SD 0.37, as shown in Table I), while Italian teachers are very close to this mean (mean 0.38, SD 0.39).[20] Interestingly, the great majority of math teachers are women and this may have important implications for the association of scientific subjects with gender. 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.

Students

Table II reports summary statistics on students' information. I restrict the sample to students with information available on the standardized test score in grade 6 and 8 and for whom I have the implicit association test of their math teacher in grade 8.[21] In the sample, 50% of students are males and boys and girls are balanced in terms of baseline characteristics related to place of birth, generation of immigration, parents' education and occupation. Test scores are standardized to have mean zero and standard deviation one per subject and year in which the

[19] Greenwald et al. (2003) suggests that a raw IAT score below -0.15 show 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.

[20] In the paper by Nosek et al. (2009), individuals completed the IAT online in the *Implicit Project* website.

[21] The Online Appendix D describes in details the sample selection and potential attrition issues.

test was taken. Females at the beginning of middle school are lagging behind of 0.19 standard deviations in math and ahead of 0.13 standard deviations in reading, with respect to males. In the same table, I also report the raw gender differences in outcomes.

The high-school track choice in this sample is comparable to national average: females are almost 10 percentage points less likely to choose an academic scientific track and almost 25 percentage points less likely to enrol in a technical technological track. Girls are more likely to choose an academic track than boys, but not a top-tier ones, which include classical and scientific tracks. Vocational school is chosen at an equal rate by both genders. However, teachers recommend 36% of males toward vocational track and 30% of females, while the scientific track is recommended only to 17% of males and 11% of females.[22]

From the original information available for a sample of students, I observe that on average there are no gender differences in assessment of ability, but females are 9 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 Italian.

4 Empirical Strategy

4.1 Estimating Equation

The main purpose of this paper is to investigate the impact of teachers' gender 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:

$$\begin{aligned} y_{ic} = \alpha_0 + \alpha_1 (Female_i \times stereotypes_c) + \alpha_2 Female_i + \eta_c + \\ + \mathbf{X}_i \rho_1 + (Female_i \times \mathbf{X}_i)\rho_2 + (Female_i \times \mathbf{Z}_c)\rho_4 + \varepsilon_{ic} \end{aligned} \tag{1}$$

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.[23] I include fixed effects at class level η_c, which absorb the average effect of teacher bias in class c. Furthermore, I include student characteristics $\mathbf{X}_i$ (initial standardized test score, parental education and occupation, immigration status and generation of immigration), and

[22] In some schools, more than one recommendation is given to students. Here, I report summary statistics only for the first recommendation.

[23] On average in 70% of the cases professors have been teaching to the same class from grade 6 to grade 8, in 11% of the cases from grade 7 and in 19% only for grade 8. Two different classes can be assigned to the same teacher.

teacher characteristics $\mathbf{Z}_c$ (as gender, place of birth, age, teacher "quality"[24], type of contract, and type of degree achieved) interacted with the gender of student i. Standard errors are robust and clustered at 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 lectures with the same classmates until grade 8. We can only identify the impact of teacher IAT score 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, α_1, measures how the gender gap in the class is affected by the assignment to teachers with one standard deviation higher stereotypes.[25] I expect the estimate of α_1 to be attenuated for the measurement error in the gender IAT score. Indeed, occasion-specific noise may introduce an attenuation bias, as suggested by Glover et al. (2017).[26] 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 differential impact on boys and girls of gender, background, and other observable characteristics of teachers. Furthermore, this allows to establish whether the impact of teachers' stereotypes on gender gap among classmates 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 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:

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

where η_{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 by school by cohort fixed effects. The advantage with respect to specification (1) is that we can analyze the impact of teachers' stereotypes separately on male students (β_3) and on female students ($\beta_1 + \beta_3$). The drawback is that I cannot

[24]Teacher "quality" is proxied by being the teacher in charge of math Olympics in the school, refresher courses and years of experience. Online Appendix Table A.V shows that being the teacher in charge of math Olympics in the school is correlated with the value added, especially for females.

[25]I discuss the exogeneity of student assignment to teachers in Section 4.3.

[26]Glover et al. (2017), while analyzing the impact on 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.

control for unobservable characteristics at the teacher or class level: this specification exploits variation in the level of teachers' stereotypes to which students of the same gender in the same school and cohort are exposed.

4.2 Correlation between implicit bias and individual characteristics

IAT scores are correlated with observable characteristics of teachers. 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 Italian) have stronger implicit association male-math and female-literature. This result suggests that individuals possess implicit gender stereotypes in self-favourable form 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 dig deeper into the determinants captured by reaction 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 does 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 35 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).[27] I further investigate this aspect by providing evidence that women labor force participation in the province of origin of teachers is negatively correlated with the IAT score (Panel B, column 2). Furthermore, I use, as proxy of cultural norms in the province of birth, the answers to the World Value Survey question on the relative rights of men and women to paid jobs when the latter are scarce.[28] I find a positive correlation between less conservative gender norms measured by this question and IAT scores (Panel B, column 3). During the survey I administered, I asked the same question to teachers themselves and I find 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

[27] 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, similarly to the average of OECD.

[28] 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".

and women and I find a weak positive correlation indistinguishable from zero. This result is not surprising in light of social psychology literature, where implicit often differ from explicit and self-reported stereotypes (Lane et al., 2007; Nosek et al., 2002).

In Panel C, columns 1 and 2, I correlate the IAT score with qualifications of the teacher (type of degree and whether the degree was achieved with honour). I find negative point estimates and high standard errors. Another rough proxy of quality of teachers is tenure (which is associated with higher experience in teaching), and being the professor in charge of math Olympiads in the school.[29] Also in these cases, point estimates are small and indistinguishable from zero. I also check whether the Gender-Science IAT score is correlated with the race IAT score. In the same regression as in Table III, I find that the correlation is -0.068 (standard error 0.123). Hence, math teachers more biased in one sphere are not more biased also in the other sphere. The IAT score does not seem to capture a general "ability" in doing this type of test for math teachers (for further information, see Alesina et al., 2018). Online Appendix Table A.IV shows jointly all correlation presented in separate regressions in Table III. Interestingly, the results are substantially invariant: gender and place of birth of teachers are the two most relevant aspects in affecting IAT scores in all specifications.

4.3 Exogeneity Assumption

Next, I present evidence on the absence of a systematic correlation between gender stereotypes of teachers and student characteristics. If parents are able to guess who is the teacher with more stereotyping behaviour, they may try to (informally) affect class assignment of their daughters. Although this seems unlikely because implicit stereotypes are not easily observable, it is also possible that they try to select teachers according to characteristics correlated with IAT score, such as gender and place of birth.[30] In column 1 of Table IV, I provide evidence that girls are not systematically assigned to teachers with more or less 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 socio-economic background – the difference is not statistically significant and the point estimate goes in the opposite direction. In

[29] In each school, usually only one professor is in charge of math Olympiad and anecdotally she is a highly motivated and passionate teacher. Indeed, as shown in the Online Appendix Table A.V, teachers in charge of math Olympics induce higher improvements in test scores of their students.

[30] In Italy, parents dislike being assigned to a teacher with a temporary contract that may have little experience and may change during the years of middle school. This paper focuses on variation of exposure to a sample of teachers that has been teaching in the same school since at least 2014. They have a lot of experience (on average 23 years) and almost all have a full-time contract. Furthermore, even if some parents manage to allocate their children to teacher with higher "quality", it does not necessarily mean that they are less gender biased. For instance, the teacher in charge of math olympics in the school is usually one of the best math teacher (as shown also in the Online Appendix Table A.V). However, if anything, they are slightly more gender biased than others (Table III).

columns 3, 4, and 5 of Table IV, I analyze the correlation respectively with father occupation, immigration background, and with a the proxy of ability using standardized test scores in math in grade 6 and I do not find statistically significant correlation. The point estimates are also small in terms of magnitude and the results are similar including all characteristics jointly (column 6). Finally, in the last column, I also include the standardized test score in math in grade 5, before entering middle school, although the sample size is substantially reduced for data availability issues.[31] The assumption of "as good as random" assignment of students to teachers with different IAT score, within a school, seems to be supported in this context. The results are identical when observations are collapsed at teacher level, as shown in the Online Appendix Table A.VI. I also check that teachers with higher bias are not systematically associated with fewer females in the top or bottom of the distribution. I find that this is not the case, as shown by the results considering the share of female students in the top 20, top 50 and bottom 20 percent of the distribution in the standardized test score in grade 6 presented in the Online Appendix Table A.VII. Finally, principals need to assign all math teachers to a class, even if he or she can guess who is the teacher with stronger stereotypes, because they don't have excess of teachers' paid hours.

The second aspect regards the absence of systematic grouping of students by socio-economic 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 socio-economic background across classes and therefore to 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, also considering that students change all teachers and school building from elementary to middle school. I check whether class assignments are statistically independent 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 7.8% of the tests performed, the p-value is lower or equal than 5%[32]. This implies that for only 7.8% of the classes we cannot reject that there is non-random assignment of one background

[31]Unfortunately, for reasons related to confidentiality, I have obtained the standardized test scores in grade 5 only for those students that did not change school code between elementary and middle school. There are only few students for which I have this information.

[32]Given the size of the Table, it is not reported in the paper but it is available upon request to the author.

characteristic. There is no evidence of systematic grouping of students according with their socio-economic background.

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 2013 and June 2015 (the detailed timeline is available in Figure II). As for Glover et al. (2017), there is a main advantage from exploiting this timeline: 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 2016.

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 an 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.[33]

Reverse causality seems unlikely for several reasons. First, as shown in Section 4.3, teachers with more stereotypes are not systematically assigned to a differential treatment in terms of student characteristics, such as family background and standardized test scores in math (see Table IV, Online Appendix Table A.VI and A.VII). Second, under the assumption of monotonic decay of the influence of exposure to students, the effect of teachers' stereotypes on student outcomes should be higher for the most recent cohort. Reverse causality does not seem to be an issue given that results are stable in all three cohorts (Online Appendix Table A.VIII). Third, math teachers included in our analysis have been teaching on average for 23 years (with a median of 25 years) and therefore over time they were exposed to hundreds of females and males students. Finally, for data availability issues[34], I do not include in the sample the cohort of student graduating right before the school year in which the IAT test was administered. Each teacher has been exposed on average to 4 classes (around 100 students) after those students included in our analysis graduated from middle school.

[33] 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 (despite being usually of less than one month in most studies) (Nosek et al., 2007).

[34] Students who were enrolled in middle school in the school year 2015-2016 and 2016-2017 are not included in the sample because they did not take one of the standardized test score.

5 The Impact of Teachers' Implicit Bias

5.1 Performance in math

By the age of 14, girls lag 0.22 standard deviations behind in math compared to their male classmates (column 1, Table V).[35] As children complete more years of education, the difference between boys and girls increases. The additional gender gap in math generated during the last two years of middle school is around 0.08 standard deviations, as shown in column 2 of Table V. This paper analyzes what happens to the gender gap when students are assigned to teachers with stronger gender stereotypes.

Table V shows the effect of teachers' implicit stereotypes on gender gap in math performance within the class, presenting the results of estimating equation (1). Classes that are assigned to teachers with one standard deviation higher IAT score have 0.03 standard deviations higher gender gap in math performance. It corresponds to an increase of 38 percent of the gender difference in performance generated during middle school, considering an average gap of 0.08 standard deviations. Column 4 includes student characteristics $(\mathbf{X}_i)$ and their interaction with gender of the children: adding these controls does not change the coefficient of interest. In the Online Appendix Table A.VIII, I show the effect of the main specification presented in Table V for the three different cohorts of students separately. Reassuringly also for the potential reverse causality concerns expressed in Section 4.3, results are not statistically different in the three cohorts.[36]

Although the level of teachers' stereotypes and all characteristics are absorbed by class fixed effect, as clarified describing equation (1), column 5 includes the interaction between student gender and teacher characteristics $(\mathbf{Z}_c)$. If anything, the coefficient of interest *(Fem*Teachers' Stereotypes)* slightly increases in magnitude when all these interaction effects are absorbed. Observable characteristics of teachers, interacted with students' gender, are not driving the relation between gender gap and teachers' stereotypes. I report the coefficients only for the main characteristics of teachers interacted with students' gender, but the effects are mainly small and insignificant for all variables, including age, parents' education, whether he or she has daughters, whether he or she achieved the degree with *laude*, the type of teaching contract, refresher

[35] This result is a result comparable to several other countries (Fryer Jr and Levitt, 2010; Bharadwaj et al., 2016). In the Online Appendix Figure A.II, 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 without controlling for class fixed effects is substantially invariant (0.20 standard deviations as shown in Table II). Most of the variation in math performance is within classes.

[36] For the first cohort, I have fewer observations because some schools change the code identifying the school that year for administrative reasons and I am not allowed to access data identified with the older codes.

courses and appointment as teacher in charge of math Olympics. The latter controls are crude proxies for teachers' value added (see the Online Appendix Table A.V).

Finally, as it can be seen in column 5 of Table V, ceteris paribus, female students assigned to female teachers have slightly lower, albeit insignificantly so, math performance in test scores in grade 8 compared to their classmates[37]. The absence of a differential impact on boys and girls of teacher gender is coherent with the result of Bharadwaj et al. (2016). However, other studies find that having a teacher of own gender helps improve performance, especially at college level (Dee, 2005; Carrell et al., 2010). In the Online Appendix Table A.IX, I split the sample among student of male and female teachers. Although the effect is not statistically significant for pupils assigned to male teachers due to a small sample size, the point estimate shows that the impact of teachers' implicit stereotypes on student performance is similar in terms of magnitude for male and female teachers. What seems to matter is whether the teacher has gender stereotypes and not the gender of the teacher per se – even if as I discussed above men tend to have higher IAT scores than women among math teachers.

To give a clearer interpretation, Figure III reports the same estimates of Table V using a categorical variable instead of the continuous one. 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. Being assigned to a teacher with a "boys-math" attitude (45% of teachers) compared to a teacher with a "girls-math" attitude (25% of teachers) leads to triple the gender gap in math improvements within the class (from -0.035 standard deviations to -0.10 standard deviations). The same results are reported in columns 1-3 in the Online Appendix Table A.X, while in columns 4-6 I consider a dummy for whether IAT score of the teachers is positive or negative, finding a similar pattern.

Are biased teachers worse instructors or are they helping boys to learn math? Next, I investigate the effect of teacher bias from estimating directly equation (2), comparing students of the same gender within the same school and cohort, but assigned to different classes. Figure IV shows that having a teacher with strong gender stereotypes 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 score. Column 3 of Table VI mirrors Figure IV: it presents the results of the regression analysis and shows that girls are lagging behind when assigned to more biased teacher, while boys are not affected by teachers' stereotypes. The results are robust to the inclusion of the

[37]It should be noticed, however, that most of teachers in Italian middle schools are females, also in math. There is little variation on the gender of teachers.

same controls as in Table V. In this specification the characteristics of teachers are not absorbed by class fixed effects and therefore controls at teacher level are particularly relevant. Column 5 is the preferred specification.

The differential response by gender is consistent with the previous results in the economic literature: women are more responsive to negative feedbacks then men in stereotypically male fields, such as math (Kugler et al., 2017). Coffman (2014) finds that individuals are significantly less likely to contribute with their ideas in gender incongruent fields and this is particularly strong for women, leading to more missed opportunities among female in male-typed categories than for males in female-typed categories. Furthermore, the type of task affects gender differences in the willingness to complete, with wider gaps in stereotypically male tasks (Niederle and Vesterlund, 2010; Große and Riener, 2010).

Heterogeneous effects

I examine which students are the most affected by teachers' stereotypes, considering their background characteristics and the time of exposure to their teachers. Table VII shows that the effect of implicit stereotypes is stronger for the most disadvantaged groups of female students. Based on the estimates in column 2, a standard deviation increase in teacher bias leads to 0.057 standard deviations higher gender gap among students with low educated mothers and of 0.029 standard deviations among students with mothers who completed at least high-school, although the difference is indistinguishable from zero at conventional levels. In the following column, I analyze the impact of teacher bias in the three terciles of the distribution of the standardized test score in grade 6. The effect is stronger for students in the lowest terciles (-0.078, with standard error 0.029, for the lowest tercile) and turning positive, but not statistically distinguishable from zero, only for students in the top of the initial ability distribution in grade 6. Finally, the effect – if anything – is slightly stronger among immigrants, even if the difference with natives is indistinguishable from zero.

Why do girls from disadvantaged backgrounds suffer the most from the interaction with biased teachers? The empirical evidence presented is coherent 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 teachers' stereotypes and among females those strongly affected have lower initial math achievements and are at higher risk of confirming the negative expectations on their group. The Online Appendix E presents a conceptual framework that illustrates how teachers' stereotypes can differentially affect effort and outcomes of students in the bottom and the top of the ability distribution.[38] One complementary explanation, coherent with the interaction theory (McConnell and Leibold, 2001), is that female students with highly educated mothers or with higher initial level of math achievement may need less interaction with their math teacher in order to avoid lagging behind with their peers. They are more likely to have both additional support to believe in their own abilities and alternative role models. Interestingly, this result is also coherent with the evidence from Tiedemann (2002): teacher perception of 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 further this aspect, I analyze the differential effect according to the "quantity" of interaction time between teacher and students. The last two columns of Table VII analyze whether there are heterogeneous effects in terms of years of exposure and hours per week.[39] Furthermore, I exploit the fact that around 20% did not have the same teacher for all three years of middle school. However, for both variables, I do not see a statistically or economically significant pattern. Most likely the impact of teacher gender stereotypes begins at a lower intensive margin and I do not have proxies of the "quality" of teacher- student interaction that would be necessary to further investigate this mechanism.

5.2 Choice of High-School Track and Teachers Recommendation

High-school track choice is the first crucial career decision in the Italian schooling system. Students and their families are free to choose their most-preferred track, with no constraints based on grades or teachers' official track recommendation. There are three main types of high-school: academic, technical, and vocational. As shown in Table II, there are substantial gender differences in the type of track selected: the preferred choice among females are academic track related to psychology, languages and art, while for males the preferred choices are academic scientific and technical technological tracks. Students in different tracks have in most cases little to no interaction during the school day. The choice of high-school is strongly correlated with university choice, as discussed in Section 2. The scientific academic path easily opens up career opportunities in STEM related fields, while the vocational choice is highly correlated with almost no tertiary education. Hence, I explore the impact of teachers' steretypes on the track choice at the end of middle school, with a focus on the choice of the scientific academic track and on the vocational track.

Table VIII, Panel A, shows that girls are 9.5 percentage points less likely than boys to attend a scientific track and equally likely to attend a vocational track (column 1 and 5, respectively).

[38] This conceptual framework is an extension of the stereotype threat model presented by Dee (2014).

[39] Around 75% of students interact with the math teacher for six hours per week, while the rest for 9 hours per week.

Teachers also recommend less girls toward a scientific track (Panel B, column 1). The gap is reduced but still substantial after controlling for the standardized test score in math. I find a close to zero and insignificant effect of teachers' stereotypes on the gender gap in scientific track choice (Panel A, columns 2-4) and in the recommendation of teachers toward this path (Panel B, columns 2-4). The inclusion of controls at student and teacher level interacted with the gender of pupils do not affect the point estimates of interest.

Recent work suggests that women are more responsive to negative feedback than men in STEM fields (Kugler et al., 2017). However, the scientific track is chosen by females with highly educated parents or with high achievement test scores, whose performance was not affected by teacher bias, as shown analyzing the heterogeneous effects in Section 5.1.[40] Girls on 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.

Teachers' stereotypes have strong impact at the bottom of the ability distribution. Indeed, as reported in columns 6 of Panel A, girls, when assigned to a teacher with one standard deviation higher implicit stereotypes, are more likely than their male classmates to attend vocational track by around 2.5 percentage points. This effect corresponds to an increase of 16% with respect to the mean probability of attending vocational training for girls. The subsequent two columns include characteristics of teachers and pupils and their interaction with the gender of the latter. Adding these controls does not change the coefficient of interest. This result mirrors an analogous differential in teachers' track recommendation toward vocational school as shown by Panel B (columns 6-8). When exposed to less gender-biased environment, female students are more likely to attend the technical track (see the Online Appendix Table A.XI).

Figure V reports the same estimates using a categorical variable instead of the continuous one to offer a clearer representation of the results.[41] Girls assigned to a teacher with a "boys-math" attitude have a probability of 18.5% of attending the vocational track, while female students assigned to a teacher with a "girls-math" attitude have a 6.1 percentage points lower probability of attending the same track, which corresponds to a decrease by 33%.

The Online Appendix Table A.XII shows the results estimating equation (2), with school instead of class fixed effects. They confirm the previous evidence of a substantial impact on female students in terms of choice of vocational training. Furthermore, boys are slightly less likely to be recommended by their teachers toward vocational track and also less likely to choose it. Finally, Online Appendix Table A.XIII presents results from the heterogeneity analysis and,

[40] In the questionnaire administered to teachers, I ask them why girls, compared to boys with the same math performance, are less likely to attend the scientific track: the reason identified as the most important by teachers is the parental influence (for the summary statistics see Table I).

[41] I use the thresholds defined by Greenwald et al. (2003) and exploited also in Figure III. For more details, check Section 5.1.

as expected, the impact of teacher bias has a stronger effect on the track choice of female students from disadvantaged background. The enrolment of females from the bottom tercile of the distribution increases by 5.1 percentage points for one standard deviation higher bias of the math teacher (which corresponds to a 16.4% increase with respect to the mean value for this group).

5.3 Additional Outcomes

Reading Performance and Stereotypes of Literature Teachers

Table A.XIV in the Online Appendix exploits the information on literature teachers that I have collected through the same IAT test and questionnaire. The gender gap in reading is reversed compared to the one in math, similarly to most OECD countries (Fryer Jr and Levitt, 2010). Female students are 0.196 standard deviation better in reading compared to male classmates at the end of the middle school and the gap is increasing of 0.082 standard deviation from grade 6 to 8, as shown in the first two columns of the Online Appendix Table A.XIV. However, the gender stereotypes of literature teachers are not affecting this gap, as shown in the subsequent columns of the same Table. The effect of interest is close to and indistinguishable from zero and it is not affected by the inclusion of teacher and student controls.

The Online Appendix Table A.XV investigates the impact of teacher stereotypes, considering the implicit IAT of both literature and math teachers and therefore restricting the sample to those classes for which these information are jointly available. The implicit stereotypes of literature teachers does not have a significant impact neither on math nor on reading standardized test scores (columns 1-4). Hence, the inclusion of their IAT score does not affect the negative and statistically significant effect of math teacher' stereotypes on math performance. Interestingly, being assigned to a math teacher with stronger implicit stereotypes seems to have negative, although indistinguishable from zero, effect on performance in reading, suggesting that female students do not simply substitute their effort in math with more effort devoted to studying literature.

Bias in Grading

Previous literature has shown the importance of gender bias in grading (i.e the gender difference in blindly graded standardized test score and teacher-assigned grades) in affecting performance in math and university choice (Lavy and Megalokonomou, 2017; Lavy and Sand, 2015; Terrier, 2015). 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 Table

A.XVI, girls on average get a higher grade compared to boys with the same standardized test score in math[42]. Females assigned to teachers with more stereotypes get a slightly lower grade, but the effect is indistinguishable from zero. However, it should be considered that grades are categorical variable from 2 to 10, where 6 is the pass grade. As it can be clearly seen by Figure A.III in the Online Appendix, there is a high bunching at the pass grade and almost half of the students obtain the same grade in math. Hence, there is little variability in teacher-assigned grades at the bottom of the distribution.

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 who attended the test score in grade 6 (9837 students), 6.0% of males and 3.3% of females are retained in (at least) one of the three years of middle school. In Table A.XVII, I check whether math teachers' bias has an impact on retention rate, but I do not find any significant impact, neither without nor with the inclusion of the controls at teacher and student level. Furthermore, I also check that teacher implicit stereotypes does not differentially impact the probability of taking the standardized test score in grade 8 (Table A.XVII, columns 5-8), conditional on taking the one in grade 6. These results suggest that the sample used in our main table on performance in math is not biased by differential attrition by gender, induced by teacher bias. Additional checks on potential sample selection issues are addressed in the Online Appendix D.

Explicit Bias

In Table A.XVIII in the Online Appendix, I consider the impact of self-reported gender differenced in innate math abilities on student outcomes. I find that it has an impact indistinguishable from zero, although the effect of more conservative gender norm on girls' performance is negative, in the same direction as the results reported by Alan et al. (2018). However, unfortunately the proxies available of gender norms are few and with little variation.[43] Finally, the impact of IAT score on student achievement is not significantly affected when I control for reported bias. This evidence seems to support the distinctiveness of implicit and explicit cognition (Greenwald

[42]Here, the results are shown controlling for the standardized test score in grade 6, but the magnitude and significance is very close when controlling for the standardized test score in grade 8.

[43]I collected also more detailed information on the reasons behind the gender gap in track choice between girls and boys who are more clearly correlated with the performance of students by gender, but in this context may be affected by reverse causality.

et al., 1998) in the context of gender stereotypes of teacher.

6 Discussion of Potential Mechanisms

In this section, I discuss the mechanisms behind the negative impact of teacher stereotypes on student achievement. I focus mainly on self-confidence and I use student survey data to provide supporting evidence. I conclude this section by discussing the channel related to the interaction theory.[44] In the Online Appendix E, I present a conceptual framework including both aspects.

Self-Confidence

Self-confidence plays a crucial role in affecting performance, especially in "gender-incongruent areas", such as math for girls (Coffman, 2014). According with social psychology, the development of academic self-concept begins since childhood and it is strongly influenced in the period after elementary school by stereotypes communicated by significant others, such as parents and teachers (Ertl et al., 2017). Girls may believe that both 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 may develop a lower self-assessment of own ability in the scientific field and potentially invest less in their STEM education. The 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).[45]

Table IX assesses the extent to which teachers' stereotypes affect one's own assessment of ability, for a sample of around 800 students for whom I collected self-confidence measures.[46] I present results for self-stereotypes in math in Panel A, in reading in Panel B and on average of all

[44] There is a third theory that could be consistent with the negative impact of teacher bias on female student math performance. According with the *animus theory*, teachers may dislike female students, treating them badly or giving them more unpleasant assignments, causing girls to dislike math. In our context, it seems unlikely that teachers assign different tasks to students by gender in terms of exams or homework. Furthermore, as shown in Section 5.3, teachers tend to favour female students in math grading, compared to blinded scores, as emerges in several other countries (Lavy and Sand, 2015; Terrier, 2015).

[45] Despite the rich literature in social psychology about stereotype threat since 1990s, only recently economists have directly analyzed 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.

[46] 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. Table available upon request to the author.

other subjects in Panel C. As shown in column 1 (Panel A), girls are 9.4 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 grade, as shown in PISA tests as well (OECD, 2015). However, girls are 5.2 percentage points more likely to consider themselves good in Italian (which corresponds to 6% percent higher probability than males), but on average both equally assess their own ability. In classes assigned to math teachers with higher bias, the gender gap in self-assessment of own ability in math increases. In particular, in classes assigned to teachers with one standard deviation higher IAT score, the gender gap in self-assessment increases by 4.9 percentage points, controlling for the test score in grade 6 as in our main specification in equation (1). Adding student and teacher level controls interacted with pupil gender do 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 achievement increases during middle school in classes assigned to a more biased teacher. Hence, in the last three columns of Table IX, 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 gap in own assessment is reduced by one third. However, teachers' stereotypes seems to have an additional impact on own assessment of math abilities, on top of performance in standardized test score, that may have detrimental effects for investment choices in education and occupation.

In Panel B and C of Tables IX , I focus on the impact of teachers' steterotypes on self-assessment respectively in Italian and all other subjects. Although girls do not improve their performance in reading when assigned to teachers with higher IAT score (see the Online Appendix Table A.XIV), they have more self-confidence in Italian. 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 Italian. There is no impact on other subjects. The effects are robust to the inclusion of controls at individual level (column 3 and 4), at teacher level (column 4), and also to the inclusion of standardized test score in grade 8.

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

Interaction Theory

A second potential mechanism is related to the *interaction theory* (McConnell and Leibold, 2001): teachers of scientific subjects with stronger implicit association between math and boys may spend less time (in terms of either quantity or quality of time) interacting with girls, especially those performing poorly. Teachers may choose to allocate more time or tailor math classes to the learning of boys and top-performing girls since they are more likely to attend a STEM track during high school. If so, I would expect higher achievement of these groups of students when exposed to a gender-biased teachers. However, as shown in Table VII and VI, I find that these groups of students are not affected by teacher stereotypes.

Unfortunately, I do not have measures of the "quality of interaction" between teachers and student by gender to directly test the interaction theory. The social psychology literature provides evidence that math teachers interact differently with male and female students. It has been shown that they believe math is more difficult for girls than equally achieving boys (Riegle-Crumb and Humphries, 2012; Tiedemann, 2002).[47] Furthermore, Sadker and Sadker (2010) document that teachers spend more time interacting with boys, while Hyde et al. (1990) suggests that math is taught as a set of computational methods to girls, while boys are encouraged to exert independence. Finally, Keller (2001) find that teachers convey their stereotyping of *mathematics as a male domain* through their classroom instruction and affect students' own association between math and gender.

Teachers' erroneous expectations may lead to a self-fulfilling prophecy: biased teachers may set a lower bar for the learning of students from sigmatized groups and they may fail to recognize their talent and not encourage them to fulfil their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983). This is coherent with the evidence presented in Table VIII on biased track recommendation to girls from teachers with stronger associations math-males and literature-females. Gender-biased interaction between pupils and teachers may be an important mechanisms behind the main results of this paper. Further research is necessary to fully address this aspect.

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: on average in

[47] Using Italian data from INVALSI, I show in the Online Appendix **??** that this perception of teachers mirrors a self-perception of students. Female students compared to boys with the same performance are more likely to believe their achievement is the result of effort and less likely to believe it is the result of ability.

OECD countries less than 5 percent of 15-years-old girls are planning to pursue a career in these fields compared to around 20 percent of boys according to 2015 PISA data. Culture and social conditioning has a strong impact on the development of skills and educational choices. This paper shows that one third of the gender gap in math performance created during middle school can be explained by teacher implicit stereotypes. Girls, especially those from disadvantaged backgrounds, are lagging behind when assigned to teachers with stronger "math-males" implicit associations. Boys, the group not ability- stigmatized in terms of math performance, are not affected by teacher stereotypes. Teachers' stereotypes influence high-school track choice, inducing more female students to attend a vocational school. Furthermore, they foster low expectations about own ability and lead to underperformance in male-typed domains. Indeed, girls are more likely to consider themselves bad in math at the end of middle school if they are assigned to a biased teacher, even controlling for their ability measured by standardized test scores. These findings are consistent with a model whereby ability-stigmatized groups underassess own ability and underperform fulfilling negative expectations about their achievements. Implicit associations can form an unintended and often an 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 bias measured by IAT score at this stage of development should not be used to make decisions about others, as hiring or firing decisions. 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 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 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 French schools, by offering alternative STEM role models (Breda et al., 2018). 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). Stereotypes about Immigrants. *Working Paper*.

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.

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. 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.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2017). Stereotypes. *Quarterly Journal of Economics*.

Bordalo, P., K. B. Coffman, N. Gennaioli, and A. Shleifer (2016). Beliefs about gender. *NBER Working Paper*.

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*.

Burns, J., L. Corno, and E. La Ferrara (2016). Interaction, Stereotypes and Performance. Evidence from South Africa. *Working Paper*.

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*.

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.

Ceci, S. J., W. M. Williams, and S. M. Barnett (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological bulletin 135*(2), 218.

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.

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.

Donders, F. (1868). *On the speed of mental processes. Translation by WG Kostor in Attention and performance II, ed. WG Koster*. North Holland.

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., 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.

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

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.

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. *Implicit measures of attitudes*, 59–102.

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 (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases. *NBER Working Paper* (w20909).

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2015). Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT). *American Economic Review 105*(5), 340–45.

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.

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.

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.

Sadker, M. and D. Sadker (2010). *Failing at fairness: How America's schools cheat girls*. Simon and Schuster.

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. (2015). Giving a little help to girls? evidence on grade discrimination and its effect on students' achievement. *Working Paper*.

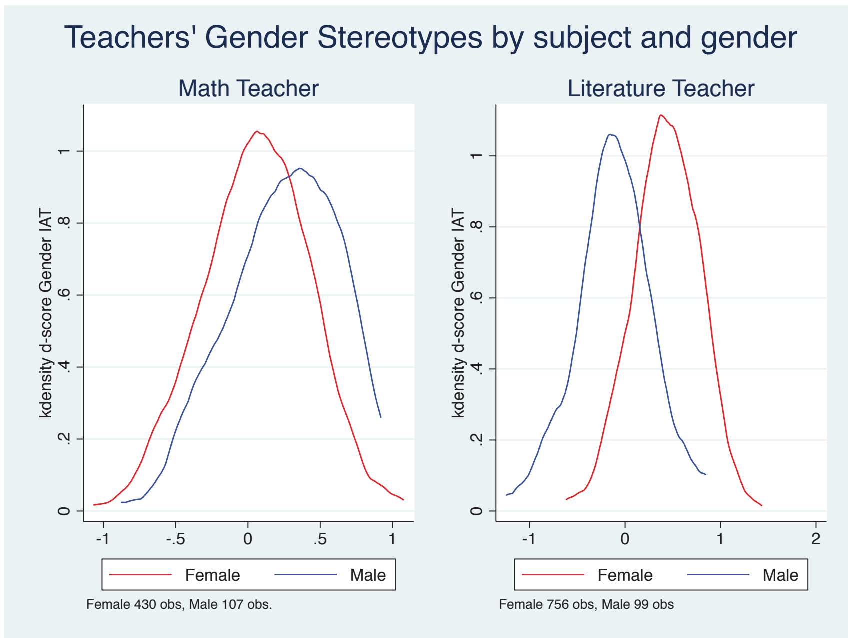
Tetlock, P. E. and G. Mitchell (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in organizational behavior 29*, 3–38.

Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics 41*(2), 191–207.

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

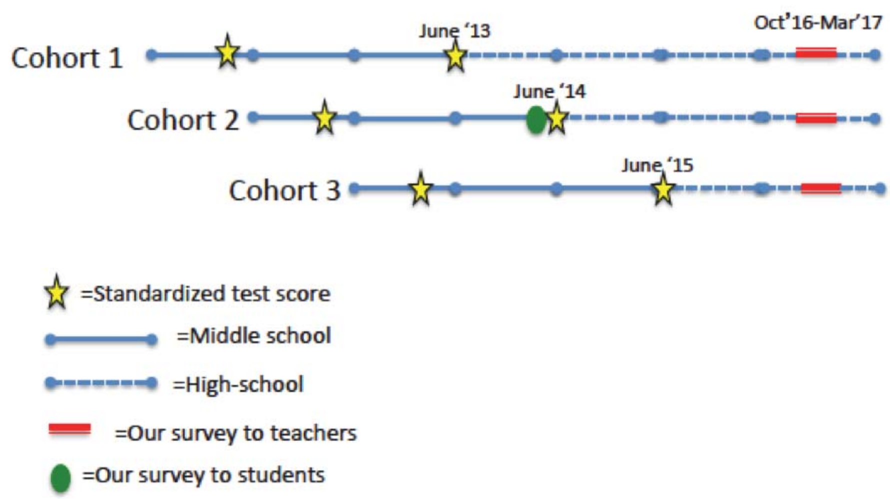
Figures and Tables

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



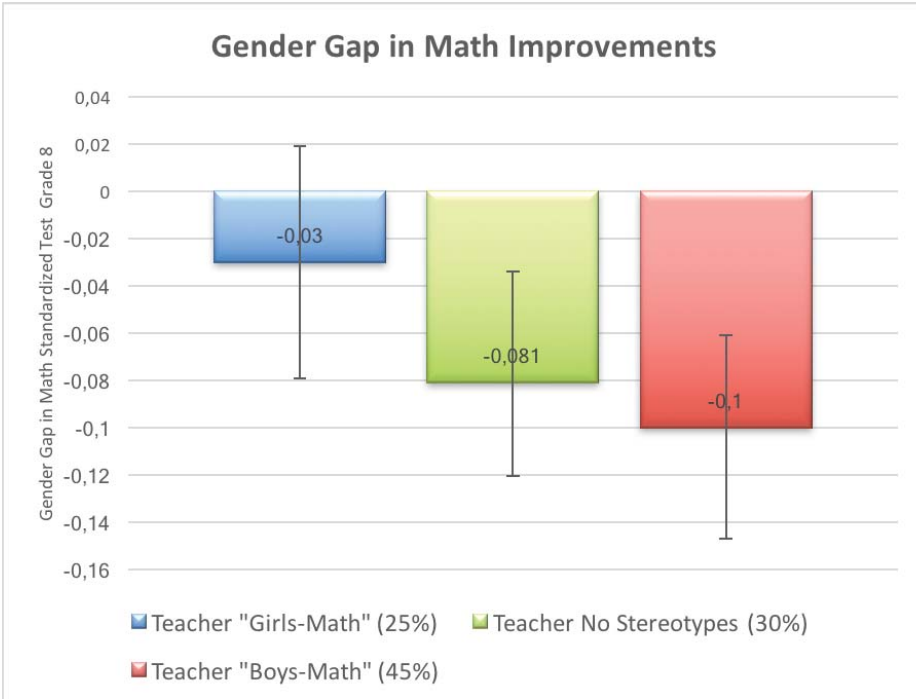
Notes: This graph shows the distribution of Gender-Science IAT scores for math and literature teachers, separated by gender. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes.

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



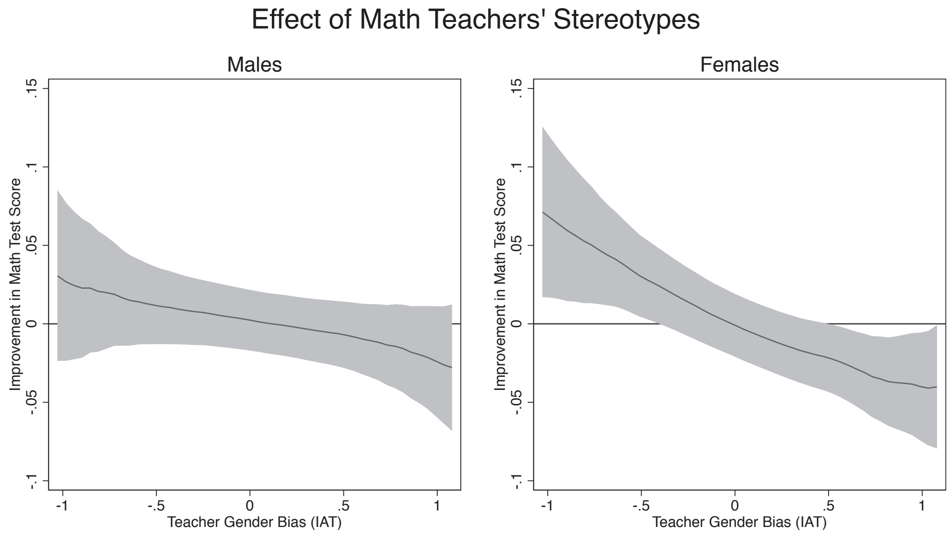
Notes: This graph shows the timeline of data collected for the three cohorts of students. They graduated from middle school between 2013 and 2015. Teachers were surveyed between October 2016 and March 2017. Standardized test scores are administered at the end of grade 6 and 8.

Figure III. Effect of teacher bias on student math performance



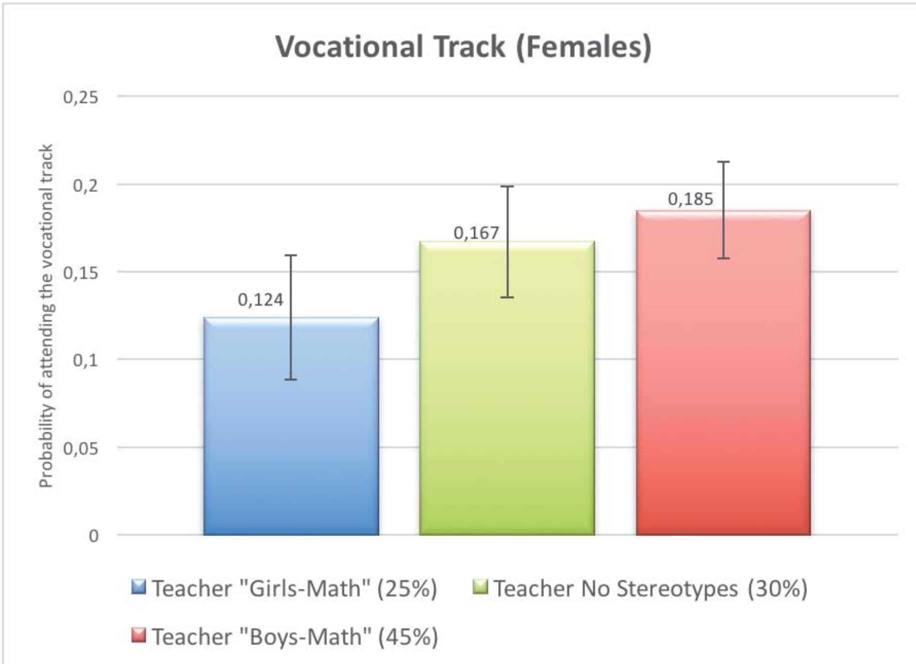
Notes: This graph shows the effect of teachers' stereotypes on student achievement. I consider the thresholds defined by Greenwald et al. (2003) where no bias is the interval of IAT raw score between -0.15 and +0.15.Teachers have a "girls-math" if the IAT score is lower than -0.15 (25% of teachers) and "boys-math" attitude if the score is higher than +0.15 (45% of teachers). The variable in the y axis is the gender gap in improvements in math between grade 6 and 8, when class fixed effects are included.

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



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

Figure V. Effect of teacher bias on choice of vocational track of females



Notes: This graph shows the effect of teachers' stereotypes on female students' track choice.I 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.Teachers have a "girls-math" if the IAT score is lower than -0.15 (25% of teachers) and "boys-math" attitude if the score is higher than +0.15 (45% of teachers).

Table I. Summary Statistics from Math Teachers' Questionnaire

	Count	Mean	SD	Min	Max
Family and education					
Female	303	0.84	0.37	0.00	1.00
Born in the North	293	0.66	0.48	0.00	1.00
Age	292	51.86	8.39	31.00	66.00
Children	303	0.75	0.44	0.00	1.00
Number of children	217	1.84	0.80	0.00	5.00
Number of daughters	217	0.86	0.77	0.00	3.00
Low edu Mother	280	0.58	0.49	0.00	1.00
Middle edu Mother	280	0.29	0.46	0.00	1.00
High edu Mother	280	0.12	0.33	0.00	1.00
Advanced STEM	294	0.24	0.43	0.00	1.00
Degree Laude	258	0.17	0.38	0.00	1.00
Job characteristics					
Full time contract	287	0.91	0.28	0.00	1.00
Years of experience	287	22.94	10.79	3.00	48.00
Math Olympiad	294	0.19	0.40	0.00	1.00
Update Courses	294	0.94	0.24	0.00	1.00
Satisfy with teacher job	290	3.70	0.84	2.00	5.00
Implicit bias					
IAT Gender	303	0.09	0.37	-1.03	1.08
Self-reported explicit bias					
WVS Gender Equality	292	0.16	0.37	0.00	1.00
Gender Dif Innate Ability	283	1.51	0.75	1.00	3.00
Reason GenderGap: Interest for STEM	258	2.57	0.98	1.00	4.00
Reason GenderGap: Predisposition for STEM	242	2.11	1.04	1.00	5.00
Reason GenderGap: Low self-esteem	279	2.64	1.05	1.00	5.00
Reason GenderGap: Family support	280	3.14	1.09	1.00	5.00
Reason GenderGap: Cultural Stereotypes	281	2.15	1.16	1.00	5.00
Boys better in Invalsi	235	0.20	0.40	0.00	1.00
Girls better in Invalsi	235	0.32	0.47	0.00	1.00
Gender Equal in Invalsi	235	0.48	0.50	0.00	1.00
Observations	303				

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 students' data is presented in Table A.III. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

Table II. Summary Statistics of students by gender

	Males	Females	Diff.	se
Baseline characteristics				
Std Math grade 6	0.230	0.036	0.194***	(0.020)
Std Ita grade 6	0.084	0.214	-0.130***	(0.019)
Born in the North	0.848	0.854	-0.006	(0.007)
Born in the Center/South	0.028	0.030	-0.003	(0.003)
Immigrant	0.189	0.174	0.015	(0.008)
Second Gen. Immigrant	0.080	0.075	0.005	(0.006)
HighEduMother	0.452	0.448	0.004	(0.010)
High Occupation Father	0.166	0.171	-0.005	(0.008)
Medium Occupation Father	0.321	0.303	0.018	(0.010)
Outcomes				
Std Math grade 8	0.190	-0.024	0.214***	(0.020)
Std Ita grade 8	-0.010	0.170	-0.180***	(0.020)
High-school Track: Scientific	0.305	0.208	0.097***	(0.010)
High-school Track: Classic	0.042	0.079	-0.036***	(0.005)
High-school Track: Other Academic	0.096	0.335	-0.239***	(0.009)
High-school Track: Technical Technological	0.311	0.067	0.244***	(0.008)
High-school Track: Technical Economic	0.113	0.163	-0.050***	(0.008)
High-school Track: Vocational	0.133	0.148	-0.015*	(0.008)
Track recommendation: Scientific	0.165	0.110	0.055***	(0.008)
Track recommendation: Vocational	0.363	0.299	0.063***	(0.011)
Own ability: all subjects	0.656	0.646	0.010	(0.012)
Own ability: math	0.833	0.747	0.087**	(0.030)
Own ability: Italian	0.917	0.968	-0.051**	(0.018)
Observations	4684	4599		

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.

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

	Panel A: Independent variables (background teachers' characteristics)				
	Female	Age	HighMotherEdu	Children	Daughters
	(1)	(2)	(3)	(4)	(5)
Dep. Var.:					
Raw IAT	-0.190**	0.009	-0.059	0.061	0.050
	(0.082)	(0.059)	(0.057)	(0.146)	(0.076)
Obs.	303	303	303	303	303
R^2	0.344	0.325	0.336	0.328	0.329
	Panel B: Independent variables (cultural traits and beliefs)				
	BornNorth	WomenLFP	WVSCityBorn	WVSIndiv	InnateAbility
	(1)	(2)	(3)	(4)	(5)
Dep. Var.:					
Raw IAT	-0.159**	-0.512**	0.391*	0.014	0.014
	(0.064)	(0.244)	(0.211)	(0.086)	(0.041)
Obs.	303	288	263	303	303
R^2	0.347	0.360	0.400	0.323	0.324
	Panel C: Independent variables (education and teacher experience)				
	STEM	Laude	FullContract	Olympiad	JobSatisfy
	(1)	(2)	(3)	(4)	(5)
Dep. Var.:					
Raw IAT	-0.094	-0.033	-0.058	0.051	0.055*
	(0.076)	(0.074)	(0.145)	(0.086)	(0.032)
Obs.	303	303	303	303	303
R^2	0.330	0.324	0.325	0.308	0.334
School FE	Yes	Yes	Yes	Yes	Yes

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 are robust and clustered at school level in parentheses; 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 is the mother of the teacher has at least a diploma,"Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. The variable "STEM" assumes value 1 if the teacher has a degree in math, engineering and physics, "Laude" is a dummy which assumes value 1 if the degree was achieved with laude, "Full Contract" assumes value 1 is the teacher has tenure, "Olympiad" is 1 for teachers in charge of math Olympiad in the school, "JobSatisfy" is a categorical variable from 1 to 5 which captures self-reported job satisfaction of teachers, "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 the latter are scarce, "WVSIndiv" is the answer to the same question at individual level, "InnateAbility" regards the teacher belief about innate differences in math abilities between men and women. 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

Dependent Variable: implicit gender stereotypes of math teacher (standardized)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fem	-0.017	-0.029	-0.021	-0.017	-0.014	-0.024	0.303
	(0.123)	(0.123)	(0.124)	(0.123)	(0.122)	(0.123)	(0.231)
Fem*HighEduMother		0.020				0.010	0.006
		(0.028)				(0.028)	(0.039)
HighEduMother		0.014				0.023	-0.015
		(0.024)				(0.023)	(0.026)
Fem*High Occupation Father			-0.006			-0.017	-0.027
			(0.035)			(0.035)	(0.054)
High Occupation Father			0.005			0.018	0.002
			(0.026)			(0.024)	(0.034)
Fem*Medium Occupation Father			0.020			0.012	0.074
			(0.030)			(0.031)	(0.050)
Medium Occupation Father			-0.011			-0.002	0.006
			(0.021)			(0.021)	(0.027)
Fem*Immigrant				-0.006		0.027	0.044
				(0.033)		(0.037)	(0.069)
Immigrant				0.055**		0.045*	0.085*
				(0.023)		(0.026)	(0.047)
Fem* Std Math grade 6					0.017	0.017	
					(0.012)	(0.012)	
Std Math grade 6					-0.016	-0.014	
					(0.011)	(0.012)	
Fem* Std Math grade 5							-0.002
							(0.018)
Std Math grade 5							-0.005
							(0.012)
School,year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	9282	9282	9282	9282	9282	9282	1649
R^2	0.484	0.484	0.484	0.484	0.484	0.485	0.765

Notes: This table reports OLS estimates of the correlation between math teachers' stereotypes measured by IAT score and students' characteristics; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 303 in columns 1-6 and 131 in column 7. The variable "Fem" indicates the gender of the student, "HighEduMother" assumes value 1 if the mother has at least a 5 years 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 manager, university professor or an executive. "Immigrant" assumes value 1 is the student is not an Italian citizen, while "Std Math grade 5" and "Std Math grade 6" are the standardized test score in grade 5 and 6 in mathematics. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for teachers' characteristics and 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 V. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

Dependent Variable: Math standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	-0.221***	-0.077***	-0.090***	-0.044	-0.038
	(0.019)	(0.014)	(0.015)	(0.033)	(0.103)
Fem*Teachers' Stereotypes			-0.030**	-0.031**	-0.042***
			(0.014)	(0.014)	(0.015)
Fem* Teacher Fem					-0.055
					(0.037)
Fem*North Teacher					0.011
					(0.030)
Fem*Advanced STEM Teacher					-0.044
					(0.031)
Std Math grade 6		0.724***	0.724***	0.699***	0.700***
		(0.012)	(0.012)	(0.013)	(0.013)
Gender Gap	-0.221	-0.077	-0.077	-0.081	-0.081
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9282	9282	9282	9282	9282
R^2	0.209	0.619	0.619	0.626	0.626

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 552. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, type of contract and education of the teacher' 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

Dependent Variable: Math standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	-0.218***	-0.077***	-0.087***	-0.027	-0.040
	(0.019)	(0.014)	(0.015)	(0.033)	(0.107)
Teachers' Stereotypes			-0.013	-0.013	-0.006
			(0.016)	(0.015)	(0.015)
Fem*Teachers' Stereotypes			-0.024*	-0.026*	-0.034**
			(0.014)	(0.014)	(0.014)
Math Teacher Fem					0.067*
					(0.040)
Fem*Math Teacher Fem					-0.050
					(0.040)
Math Teacher born North					0.031
					(0.035)
Fem*North Math Teacher					0.014
					(0.031)
Advanced STEM					0.028
					(0.034)
Fem*Advanced STEM Teacher					-0.032
					(0.034)
Std Math grade 6		0.717***	0.716***	0.689***	0.689***
		(0.011)	(0.011)	(0.012)	(0.012)
Gender Gap	-0.218	-0.072	-0.072	-0.082	-0.083
School, year FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9282	9282	9282	9282	9282
R^2	0.135	0.575	0.576	0.584	0.588

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (school by cohort) is 185. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table VII. 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	-0.038	-0.037	0.001	-0.036	-0.064	-0.056
	(0.103)	(0.103)	(0.113)	(0.103)	(0.103)	(0.106)
Fem*Teachers' Stereotypes(IAT)	-0.042***	-0.057***	-0.078***	-0.040**	-0.047***	-0.065*
	(0.015)	(0.022)	(0.029)	(0.017)	(0.018)	(0.034)
Fem*IAT*HighEduM		0.028				
		(0.030)				
Fem*IAT*Top tercile Math6			0.112***			
			(0.039)			
Fem*IAT*Middle tercile Math6			0.009			
			(0.038)			
Fem*IAT*Immigrant				-0.014		
				(0.042)		
Fem*IAT*Extended School Day					0.023	
					(0.028)	
Fem*IAT*Same Math Teacher						0.028
						(0.038)
Gender Gap	-0.082	-0.082	-0.082	-0.082	-0.082	-0.082
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.	9282	9282	9282	9282	9282	9282
R^2	0.626	0.627	0.628	0.627	0.627	0.627

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 552. The variable "Fem" indicates the gender of the student, "HighEduM" wether the mother has at least a diploma, "tercile Math6" is the tercile of standardized test score in math in grade 6 and "Immigrant" is a dummy equal to 1 if 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, type of contract and education of the teacher' 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 VIII. Estimation of the effect of teachers' gender bias on track choice- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A- Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Female	-0.095***	-0.044***	0.030	0.166*	0.014	0.001	0.028	0.053
	(0.012)	(0.012)	(0.020)	(0.092)	(0.009)	(0.011)	(0.023)	(0.073)
Fem*Stereotypes		0.011	0.010	0.003		0.025**	0.022**	0.024**
		(0.013)	(0.012)	(0.012)		(0.010)	(0.010)	(0.010)
Fem* Teacher Fem				-0.033				0.037*
				(0.029)				(0.022)
Std Math grade 6		0.178***	0.159***	0.160***		-0.104***	-0.091***	-0.091***
		(0.008)	(0.008)	(0.008)		(0.007)	(0.007)	(0.007)
Constant	0.300***	0.243***	0.107***	0.109***	0.141***	0.174***	0.206***	0.205***
	(0.006)	(0.006)	(0.015)	(0.015)	(0.005)	(0.006)	(0.016)	(0.016)
Mean Y Fem	0.205	0.205	0.205	0.205	0.155	0.155	0.155	0.155
Obs.	8451	8451	8451	8451	8451	8451	8451	8451
R^2	0.115	0.216	0.236	0.238	0.119	0.190	0.208	0.211
	Panel B- Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.045***	-0.018*	0.034**	0.028	-0.059***	-0.103***	-0.116***	-0.031
	(0.010)	(0.010)	(0.016)	(0.081)	(0.013)	(0.012)	(0.024)	(0.091)
Fem*Stereotypes		0.004	0.003	-0.005		0.018	0.018	0.024**
		(0.010)	(0.010)	(0.010)		(0.011)	(0.011)	(0.012)
Fem* Teacher Fem				-0.054**				0.025
				(0.025)				(0.037)
Std Math grade 6		0.127***	0.113***	0.114***		-0.245***	-0.216***	-0.217***
		(0.009)	(0.009)	(0.009)		(0.008)	(0.008)	(0.008)
Constant	0.156***	0.130***	0.060***	0.096	0.377***	0.429***	0.515***	2.344***
	(0.005)	(0.004)	(0.011)	(0.134)	(0.006)	(0.006)	(0.017)	(0.117)
Mean Y Fem	0.109	0.109	0.109	0.109	0.318	0.318	0.318	0.318
Obs.	7084	7084	7084	7084	7084	7084	7084	7084
R^2	0.156	0.243	0.253	0.255	0.148	0.360	0.387	0.388
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teach Controls	No	No	No	Yes	No	No	No	Yes

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 551. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)							
Fem	-0.094***	-0.086***	-0.114*	0.147	-0.066**	-0.088	0.174
	(0.029)	(0.029)	(0.066)	(0.189)	(0.028)	(0.066)	(0.200)
Fem*Teacher Stereotypes		-0.049**	-0.053**	-0.072**	-0.032	-0.036	-0.056*
		(0.023)	(0.024)	(0.033)	(0.023)	(0.025)	(0.032)
Constant	0.837***	0.808***	0.809***	0.800***	0.810***	0.820***	0.812***
	(0.015)	(0.015)	(0.048)	(0.047)	(0.015)	(0.048)	(0.046)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	747	747	747	747	747	747	747
R^2	0.110	0.216	0.236	0.253	0.248	0.266	0.281
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Fem	0.052**	0.072***	0.061	0.177	0.064***	0.051	0.148
	(0.023)	(0.025)	(0.047)	(0.214)	(0.023)	(0.045)	(0.202)
Fem*Teacher Stereotypes		0.041**	0.041**	0.029	0.042**	0.042**	0.032
		(0.019)	(0.020)	(0.024)	(0.018)	(0.020)	(0.023)
Constant	0.916***	0.908***	0.937***	0.946***	0.917***	0.953***	0.963***
	(0.012)	(0.012)	(0.034)	(0.035)	(0.011)	(0.034)	(0.035)
Std Test score Italian	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	664	664	664	664	664	664	664
R^2	0.115	0.134	0.148	0.175	0.148	0.161	0.189
Panel C- Dependent Variable: Average own ability in other subjects							
Fem	0.035	0.013	0.015	-0.222	0.008	0.009	-0.230
	(0.027)	(0.028)	(0.061)	(0.222)	(0.027)	(0.061)	(0.223)
Fem*Teacher Stereotypes		-0.015	-0.016	-0.022	-0.020	-0.021	-0.027
		(0.025)	(0.026)	(0.029)	(0.026)	(0.026)	(0.029)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	802	802	802	802	802	802	802
R^2	0.096	0.125	0.137	0.157	0.130	0.141	0.161
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	Yes	Yes
Math Teacher Controls	No	No	No	Yes	No	No	Yes

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 are robust and clustered at math teacher level in parentheses; the number of clusters for math is 58. The number of fixed effects (classes) is 62. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Online Appendix for
Implicit Stereotypes: Evidence from Teachers' Gender Bias

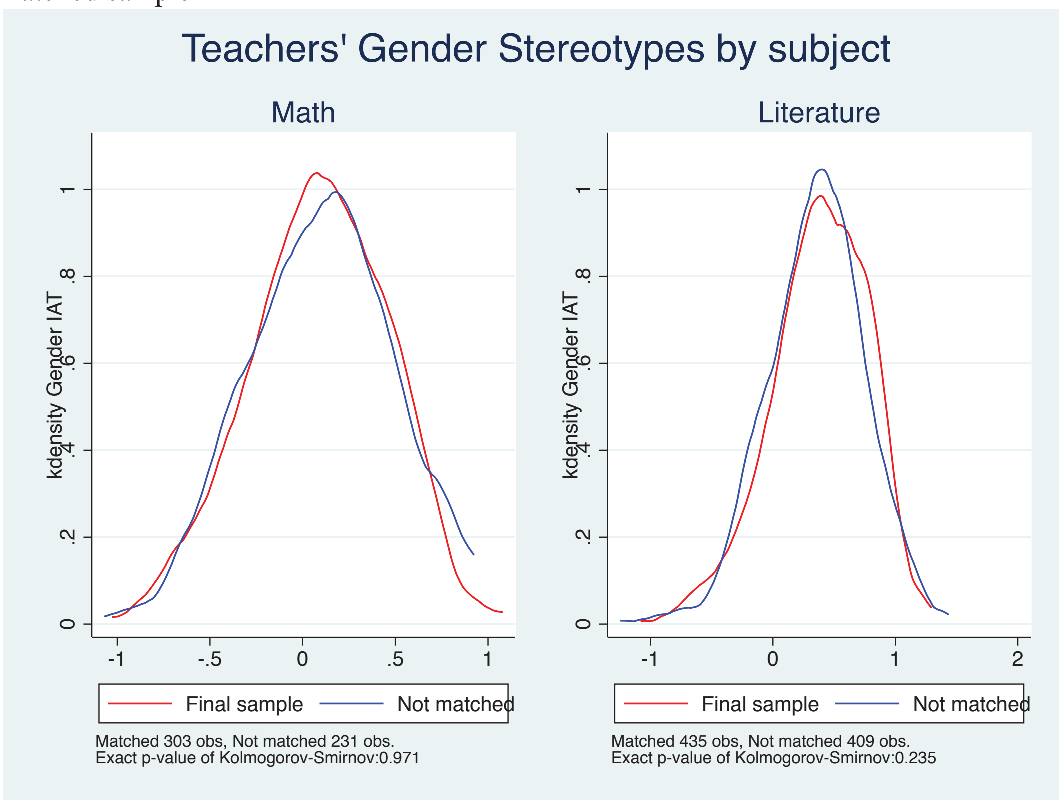
Michela Carlana*

July 12, 2018

*Harvard Kennedy School (e-mail: michela_carlana@hks.harvard.edu).

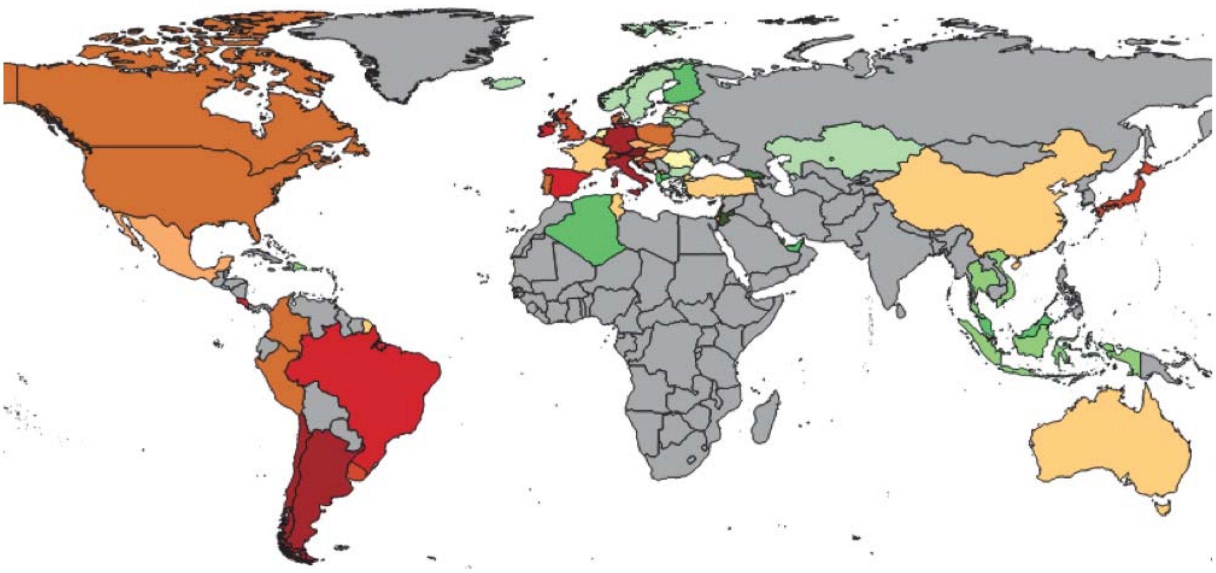
A Additional Figures and Tables

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



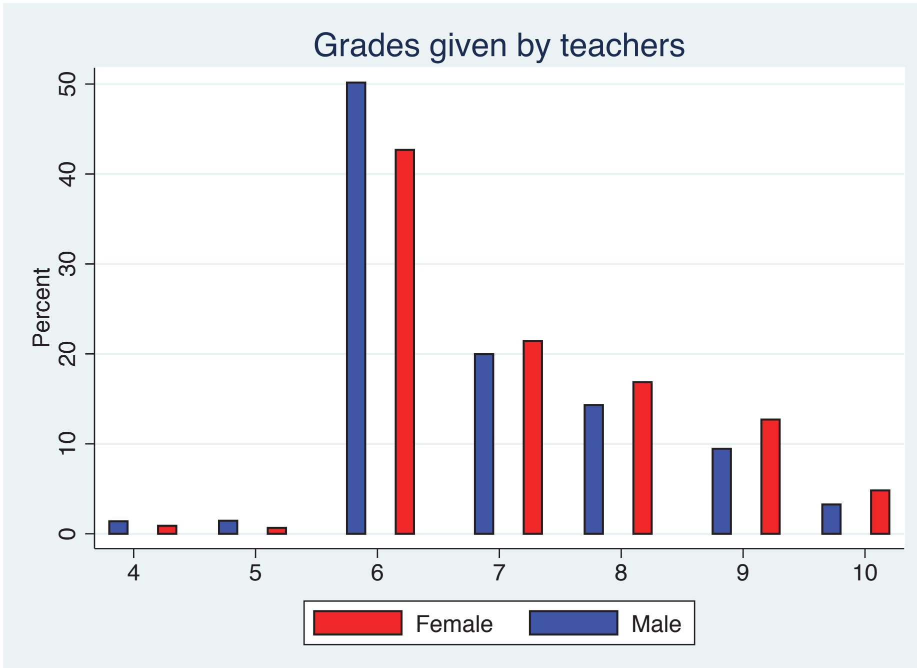
Notes: This graph shows the distribution of Gender-Science IAT scores for math and literature teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes. The graph provides evidence that data on teachers used in this papers are similar in terms of gender stereotypes compared 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 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.III: 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

Panel A: Comparison between students in the sample and other Italian students				
	(1)	(2)	(3)	(4)
Variable	Other Italian Students	Sample	Diff.	Stand. Diff.
Female	0.493	0.495	0.003	0.004
	(0.500)	(0.500)	(0.006)	
Immigrant	0.100	0.202	0.102	0.204
	(0.300)	(0.401)	(0.012)***	
Test score math grade 8 (girls)	56.884	57.580	0.696	0.027
	(18.360)	(17.789)	(0.762)	
Test score math grade 8 (boys)	59.360	61.456	2.095	0.081
	(18.490)	(18.054)	(0.814)**	
Mother: Less than Diploma	0.297	0.239	-0.058	-0.093
	(0.457)	(0.426)	(0.018)***	
Mother: Diploma	0.384	0.395	0.010	0.015
	(0.486)	(0.489)	(0.021)	
Mother: More than Diploma	0.104	0.132	0.028	0.062
	(0.306)	(0.339)	(0.021)	
Father: Less than Diploma	0.338	0.272	-0.066	-0.102
	(0.473)	(0.445)	(0.019)***	
Father: Diploma	0.338	0.352	0.014	0.021
	(0.473)	(0.478)	(0.019)	
Father: More than Diploma	0.095	0.127	0.033	0.074
	(0.293)	(0.333)	(0.021)	
Mother: Low Occupation	0.459	0.398	-0.061	-0.087
	(0.498)	(0.490)	(0.023)***	
Mother: Intermediate Occupation	0.256	0.307	0.052	0.082
	(0.436)	(0.461)	(0.017)***	
Mother: High Occupation	0.081	0.094	0.013	0.033
	(0.272)	(0.292)	(0.013)	
Father: Low Occupation	0.285	0.301	0.016	0.025
	(0.451)	(0.459)	(0.020)	
Father: Intermediate Occupation	0.314	0.307	-0.007	-0.011
	(0.464)	(0.461)	(0.018)	
Father: High Occupation	0.166	0.166	0.000	0.000
	(0.372)	(0.372)	(0.020)	
Class size	22.243	22.650	0.407	0.089
	(3.824)	(2.549)	(0.196)**	
Observations	1,558,592	9,282	1,567,874	

Notes: data from the standardized test score INVALSI of all Italian students in grade 8 from 2011-12 to 2013-14.

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

Panel B: Comparison between students in the sample and other students in the same provinces				
	(1)	(2)	(3)	(4)
Variable	Other Italian Students	Sample	Diff.	Stand. Diff.
Female	0.492	0.495	0.003	0.004
	(0.500)	(0.500)	(0.006)	
Immigrant	0.143	0.202	0.059	0.111
	(0.350)	(0.401)	(0.011)***	
Test score grade 8 (girls)	56.417	57.580	1.163	0.046
	(18.295)	(17.789)	(0.749)	
Test score grade 8 (boys)	60.006	61.456	1.450	0.056
	(18.712)	(18.054)	(0.801)*	
Mother: Less than Diploma	0.240	0.239	-0.001	-0.001
	(0.427)	(0.426)	(0.018)	
Mother: Diploma	0.425	0.395	-0.030	-0.043
	(0.494)	(0.489)	(0.021)	
Mother: More than Diploma	0.130	0.132	0.002	0.005
	(0.336)	(0.339)	(0.020)	
Father: Less than Diploma	0.288	0.272	-0.016	-0.025
	(0.453)	(0.445)	(0.019)	
Father: Diploma	0.371	0.352	-0.018	-0.027
	(0.483)	(0.478)	(0.019)	
Father: More than Diploma	0.121	0.127	0.006	0.014
	(0.326)	(0.333)	(0.021)	
Mother: Low Occupation	0.385	0.398	0.013	0.019
	(0.487)	(0.490)	(0.022)	
Mother: Intermediate Occupation	0.316	0.307	-0.009	-0.014
	(0.465)	(0.461)	(0.017)	
Mother: High Occupation	0.107	0.094	-0.013	-0.030
	(0.309)	(0.292)	(0.013)	
Father: Low Occupation	0.268	0.301	0.033	0.052
	(0.443)	(0.459)	(0.020)*	
Father: Intermediate Occupation	0.323	0.307	-0.015	-0.023
	(0.468)	(0.461)	(0.017)	
Father: High Occupation	0.195	0.166	-0.029	-0.053
	(0.396)	(0.372)	(0.020)	
Class size	22.593	22.650	0.057	0.014
	(3.128)	(2.549)	(0.199)	
Observations	217,910	9,282	227,192	

Notes: data from the standardized test score INVALSI of all Italian students in grade 8 from 2011-12 to 2013-14.

Table A.II: 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	-0.035			-0.034	-0.040
	(0.037)			(0.037)	(0.047)
First Questionnaire, then IAT		-0.151		-0.166	0.017
		(0.127)		(0.126)	(0.185)
Order Compatible IAT Gender			-0.051*	-0.052*	-0.049
			(0.029)	(0.030)	(0.039)
Constant	0.106***	0.097***	0.119***	0.134***	0.131***
	(0.019)	(0.016)	(0.020)	(0.022)	(0.024)
School FE	No	No	No	No	Yes
Obs.	534	534	534	534	534
R^2	0.002	0.002	0.005	0.009	0.168

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 it was asked to associate together first more likely compatible categories (Male-Scientific vs. Female-Humanistic) or the opposite (Female-Scientific vs. Male-Humanistic). Finally, in 8 cases for the math teacher, we asked to complete first a questionnaire and then the IATs. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.III: Balance table of the differences between teachers matched (not matched) with students graduating from 2013 to 2015

	Math Teachers				Italian Teachers			
	Not matched	Matched	Dif.	se	Not matched	Matched	Dif.	se
Female	0.761	0.838	-0.078*	(0.034)	0.859	0.908	-0.049*	(0.022)
Born in the North	0.541	0.655	-0.115**	(0.043)	0.605	0.737	-0.132***	(0.033)
Age	42.538	51.863	-9.325***	(0.768)	42.917	51.260	-8.343***	(0.596)
Full time contract	0.516	0.913	-0.397***	(0.035)	0.731	0.986	-0.255***	(0.022)
Yeas of experience	10.694	22.861	-12.167***	(0.942)	13.899	23.561	-9.662***	(0.695)
Teaching in 2015-16	0.560	0.983	-0.423***	(0.030)	0.619	0.995	-0.377***	(0.024)
Children	0.568	0.746	-0.177***	(0.040)	0.567	0.729	-0.162***	(0.032)
Number of children	1.856	1.843	0.013	(0.095)	1.670	1.788	-0.118	(0.074)
Number of daughters	0.880	0.857	0.023	(0.087)	0.920	0.814	0.106	(0.068)
Low edu Mother	0.410	0.582	-0.173***	(0.045)	0.391	0.510	-0.119**	(0.036)
Middle edu Mother	0.400	0.293	0.107*	(0.043)	0.402	0.378	0.025	(0.036)
High edu Mother	0.190	0.125	0.065*	(0.033)	0.207	0.112	0.095***	(0.026)
Advanced STEM	0.239	0.238	0.001	(0.038)	0.000	0.000	0.000	(0.000)
Math Olympiad	0.084	0.194	-0.110***	(0.031)	0.000	0.000	0.000	(0.000)
Update Courses	0.850	0.939	-0.089***	(0.026)	0.862	0.932	-0.071***	(0.021)
Degree Laude	0.273	0.171	0.103**	(0.039)	0.388	0.291	0.097**	(0.035)
IAT Gender	0.102	0.089	0.013	(0.033)	0.350	0.382	-0.031	(0.027)
Boys better in Invalsi	0.237	0.204	0.033	(0.043)	0.096	0.098	-0.002	(0.026)
Girls better in Invalsi	0.308	0.319	-0.011	(0.048)	0.546	0.530	0.016	(0.043)
Gender Equal in Invalsi	0.455	0.477	-0.021	(0.052)	0.358	0.372	-0.014	(0.042)
Satisfy with teacher job	3.741	3.698	0.043	(0.079)	3.861	3.891	-0.030	(0.061)
WVS Gender Equality	0.152	0.164	-0.013	(0.033)	0.111	0.104	0.007	(0.022)
Gender Dif Innate Ability	1.561	1.512	0.048	(0.069)	1.487	1.374	0.113*	(0.049)
Reason Gender Gap:								
Interest for STEM	2.554	2.571	-0.017	(0.095)	2.874	2.646	0.228**	(0.076)
Predisposition for STEM	2.144	2.112	0.032	(0.103)	2.227	2.154	0.074	(0.081)
Low self-esteem	2.909	2.636	0.273**	(0.094)	2.691	2.544	0.146	(0.079)
Family support	3.165	3.136	0.029	(0.098)	3.121	2.939	0.182*	(0.076)
Cultural Stereotypes	2.465	2.146	0.320**	(0.105)	2.313	2.170	0.143	(0.086)
Observations	234	303			418	435		

Notes: First hand data from teachers' questionnaire. We compare teachers' matched with students' data with those not matched. The main reason for not being able to match students with teachers are two: teachers may have started to teach in the school after Summer 2016 or I did not obtain the authorization to use the INVALSI data.

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

Dependent variable: raw IAT score of math teachers

	Restricted sample		**All teachers**	
	(1)	(2)	(3)	(4)
Female	-0.160***	-0.190**	-0.185***	-0.169***
	(0.056)	(0.089)	(0.045)	(0.060)
Born in the North	-0.055	-0.134*	-0.070**	-0.115**
	(0.048)	(0.074)	(0.034)	(0.046)
Age	0.036	0.043	-0.018	-0.035
	(0.038)	(0.061)	(0.018)	(0.024)
Age sq.	-0.000	-0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
High Edu Mother	-0.011	-0.013	0.009	0.002
	(0.044)	(0.058)	(0.036)	(0.044)
Children	0.197	0.026	0.021	0.147
	(0.125)	(0.096)	(0.054)	(0.338)
Daughters	0.077	0.050	0.017	-0.015
	(0.065)	(0.088)	(0.049)	(0.059)
Advanced STEM	-0.071	-0.103	-0.091**	-0.135***
	(0.053)	(0.078)	(0.042)	(0.051)
Degree Laude	-0.067	-0.044	-0.064	-0.053
	(0.059)	(0.082)	(0.042)	(0.050)
Full time contract	-0.071	-0.083	0.011	0.002
	(0.086)	(0.130)	(0.056)	(0.075)
Math Olympiad	0.093	0.070	0.078	0.102
	(0.067)	(0.091)	(0.054)	(0.071)
Satisfy with teacher job	0.039	0.043	0.012	0.015
	(0.025)	(0.035)	(0.018)	(0.020)
WVS Gender Equality	0.023	0.005	-0.007	-0.018
	(0.061)	(0.094)	(0.047)	(0.064)
Gender Dif Innate Ability	0.018	0.026	0.025	0.030
	(0.029)	(0.045)	(0.020)	(0.029)
School FE	No	Yes	No	Yes
Obs.	303	303	534	534
R^2	0.120	0.423	0.094	0.281

Notes: This table reports OLS estimates of the correlation between math teachers' stereotypes measured by IAT score and own teacher characteristics; the unit of observation is teacher t in school s. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90. We include the order of IATs for math teachers and missing categories if the information is not available. The restricted sample includes data exploits on teachers in the main regressions of this paper. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.V: The impact of math teacher characteristics on students' improvement in performance

Dependent variable: Std Math test score in grade 8									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: X=	**Teacher Female**			**Teacher born North**			**Advanced STEM**		
X	0.079***	0.085**	0.082*	-0.006	-0.020	0.013	0.008	0.016	0.028
	(0.030)	(0.037)	(0.043)	(0.031)	(0.035)	(0.030)	(0.032)	(0.035)	(0.033)
Fem*X		-0.013	-0.031		0.029	0.020		-0.015	-0.009
		(0.039)	(0.038)		(0.030)	(0.028)		(0.033)	(0.031)
Obs.	9283	9283	9282	9283	9283	9282	9283	9283	9282
R^2	0.528	0.528	0.562	0.527	0.527	0.562	0.527	0.527	0.562
Panel A: X=	**Math Olympics**			**Full-time contract**			**Degree with laude**		
X	0.077**	0.039	0.028	0.046	0.029	0.020	0.042	0.045	0.073**
	(0.039)	(0.045)	(0.040)	(0.043)	(0.050)	(0.047)	(0.039)	(0.044)	(0.035)
Fem*X		0.077*	0.074*		0.033	0.023		-0.004	-0.002
		(0.042)	(0.038)		(0.073)	(0.066)		(0.041)	(0.040)
Obs.	9283	9283	9282	9283	9283	9282	9283	9283	9282
R^2	0.528	0.529	0.563	0.527	0.527	0.562	0.527	0.527	0.563
School FE	No	No	Yes	No	No	Yes	No	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 are robust and clustered at teacher level in parentheses. All columns include a dummy for the gender of the student ("Fem"), standardized test score in grade 6 and the interaction with the gender of the student. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

Dependent Variable: Math Teacher implicit gender bias (standardized)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share Fem	0.740					0.458	0.602	-1.347
	(0.896)					(0.875)	(0.795)	(3.273)
Share Male HighEduMother		0.145				0.146	0.232	-0.031
		(0.574)				(0.583)	(0.599)	(1.646)
Share Fem HighEduMother		0.377				0.159	0.118	1.361
		(0.517)				(0.520)	(0.574)	(1.769)
Share Male HighOccFather			0.752			1.260	1.279	0.957
			(0.878)			(0.898)	(0.860)	(3.195)
Share Fem HighOccFather			0.007			-0.417	-0.514	0.029
			(0.607)			(0.716)	(0.815)	(1.907)
Share Male MedOccFather			0.033			0.523	0.332	0.327
			(0.618)			(0.634)	(0.662)	(1.138)
Share Fem MedOccFather			0.743			0.669	0.322	0.437
			(0.683)			(0.664)	(0.715)	(1.643)
Share Male Immigrant				0.859		1.175**	1.127*	1.298
				(0.592)		(0.568)	(0.578)	(1.363)
Share Fem Immigrant				-0.131		-0.063	-0.002	-0.594
				(0.563)		(0.588)	(0.645)	(1.312)
Male Average Std Math 6					-0.032	-0.104	-0.181	
					(0.196)	(0.199)	(0.209)	
Female Average Std Math 6					0.169	0.190	0.248	
					(0.221)	(0.242)	(0.284)	
Male Average Std Math 5								0.156
								(0.386)
Female Average Std Math 5								-0.251
								(0.539)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	No	Yes	Yes
Obs.	303	303	303	303	303	303	303	109
R^2	0.327	0.342	0.349	0.337	0.325	0.388	0.473	0.682

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and students' characteristics; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90 in columns 1-6 and 40 in column 7. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression 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.VII: Correlation between IAT score of teacher and share of females among best performing students

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: raw IAT score math teachers						
Share Females Top20	0.048	0.397				
	(0.242)	(0.356)				
Share Females Top50			0.341	0.576		
			(0.403)	(0.521)		
Share Females Bottom20					0.082	0.197
					(0.215)	(0.316)
School FE	No	Yes	No	Yes	No	Yes
Teacher Controls	No	Yes	No	Yes	No	Yes
Obs.	303	303	303	303	303	303
R^2	0.012	0.422	0.015	0.422	0.013	0.419

Notes: This table reports OLS estimates of the correlation between math teacher bias, measured by IAT score, and share of top and bottom performing females in grade 6; the unit of observation is teacher t. Standard errors are robust and clustered at school level in parentheses. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

	Dependent Variable: Math standardized test score in grade 8			
	All Students	First Cohort	Second Cohort	Third Cohort
	(1)	(2)	(3)	(4)
Fem*Teachers' Stereotypes	-0.042***	-0.049	-0.038	-0.044**
	(0.015)	(0.043)	(0.025)	(0.021)
Std Math grade 6	0.700***	0.671***	0.692***	0.722***
	(0.013)	(0.032)	(0.021)	(0.016)
Class FE	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
Obs.	9282	1984	4116	3182
R^2	0.626	0.623	0.609	0.661

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 552, 120, 250, 182, respectively in the four columns. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

Dep. Variable	**Std Math Grade 8**					
	All		**Female Teachers**		**Male Teachers**	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.090***	-0.044	-0.097***	-0.063*	-0.050	0.043
	(0.015)	(0.033)	(0.017)	(0.036)	(0.038)	(0.084)
Fem*Teachers' Stereotypes	-0.030**	-0.031**	-0.033**	-0.034**	-0.033	-0.027
	(0.014)	(0.014)	(0.015)	(0.015)	(0.044)	(0.044)
Std Math grade 6	0.724***	0.699***	0.726***	0.701***	0.714***	0.691***
	(0.012)	(0.013)	(0.013)	(0.013)	(0.034)	(0.036)
Student controls	No	Yes	No	Yes	No	Yes
Obs.	9282	9282	7976	7976	1306	1306
R^2	0.619	0.626	0.621	0.628	0.602	0.612

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 are robust and clustered at math teacher level in parentheses. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

	Std Math 8th grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.030	0.020	0.058	-0.044**	0.004	0.019
	(0.025)	(0.036)	(0.103)	(0.020)	(0.035)	(0.102)
Fem *No Bias	-0.051	-0.054	-0.049			
	(0.034)	(0.034)	(0.033)			
Fem *Male-Math	-0.068**	-0.074**	-0.095***			
	(0.032)	(0.032)	(0.032)			
Fem *IAT>0 Male-Math				-0.055**	-0.057**	-0.073***
				(0.026)	(0.026)	(0.027)
Std Math grade 6	0.724***	0.699***	0.700***	0.725***	0.699***	0.701***
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Constant	0.025***	-0.116***	-1.230***	0.025***	-0.116***	-1.215***
	(0.007)	(0.023)	(0.152)	(0.007)	(0.023)	(0.152)
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.	9282	9282	9282	9282	9282	9282
R^2	0.619	0.626	0.626	0.619	0.626	0.626

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 we use categorical variables. In columns 1-3, 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. In columns 4-6, we consider a positive or negative sign in the IAT score. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, type of contract and education of the teacher' 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 technical technological track

Dependent Variable: Track choice Technical Technological				
	(1)	(2)	(3)	(4)
Fem	-0.244***	-0.266***	-0.352***	-0.387***
	(0.011)	(0.012)	(0.024)	(0.086)
Fem*Teachers' Stereotypes		-0.023*	-0.020	-0.015
		(0.013)	(0.012)	(0.012)
Fem*Math Teacher Fem				0.015
				(0.036)
Fem*North Math Teacher				0.006
				(0.023)
Fem*Advanced STEM Teacher				0.020
				(0.026)
Std Math grade 6		-0.041***	-0.029***	-0.030***
		(0.007)	(0.007)	(0.008)
Constant	0.311***	0.324***	0.409***	0.407***
	(0.006)	(0.006)	(0.018)	(0.018)
Mean Y for Fem	0.07	0.07	0.07	0.07
Class FE	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes
Obs.	8451	8451	8451	8451
R^2	0.200	0.205	0.218	0.220

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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 551. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XII: Estimation of the effet of teachers' gender stereotypes on track choice- school FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.093***	-0.045***	0.027	0.151*	0.004	-0.007	0.020	0.042
	(0.014)	(0.014)	(0.020)	(0.089)	(0.010)	(0.012)	(0.023)	(0.069)
Fem*Stereotypes		0.011	0.009	0.004		0.027***	0.024**	0.026***
		(0.012)	(0.011)	(0.012)		(0.009)	(0.010)	(0.010)
Stereotypes		-0.008	-0.008	-0.006		-0.016**	-0.014**	-0.012
		(0.009)	(0.009)	(0.009)		(0.007)	(0.007)	(0.008)
Std Math grade 6		0.172***	0.151***	0.153***		-0.103***	-0.090***	-0.090***
		(0.007)	(0.008)	(0.008)		(0.006)	(0.006)	(0.006)
Constant	0.302***	0.244***	0.106***	0.021	0.140***	0.167***	0.203***	0.217***
	(0.011)	(0.011)	(0.015)	(0.075)	(0.008)	(0.009)	(0.017)	(0.053)
Obs.	8451	8451	8451	8451	8451	8451	8451	8451
R^2	0.068	0.172	0.194	0.198	0.067	0.144	0.164	0.169
	Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.058***	-0.029**	0.022	0.013	-0.058***	-0.103***	-0.116***	-0.056
	(0.012)	(0.012)	(0.016)	(0.078)	(0.015)	(0.013)	(0.023)	(0.093)
Fem*Stereotypes		0.003	0.002	-0.005		0.018*	0.019*	0.024**
		(0.010)	(0.010)	(0.010)		(0.011)	(0.011)	(0.011)
Stereotypes		-0.008	-0.008	-0.003		-0.015	-0.015*	-0.016*
		(0.008)	(0.008)	(0.008)		(0.009)	(0.009)	(0.009)
Std Math grade 6		0.124***	0.111***	0.112***		-0.237***	-0.208***	-0.209***
		(0.009)	(0.008)	(0.008)		(0.008)	(0.008)	(0.008)
Constant	0.172***	0.144***	0.073***	0.044	0.368***	0.411***	0.504***	0.464***
	(0.011)	(0.011)	(0.014)	(0.068)	(0.015)	(0.012)	(0.019)	(0.081)
Obs.	7084	7084	7084	7084	7084	7084	7084	7084
R^2	0.098	0.188	0.199	0.204	0.101	0.320	0.350	0.353
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. 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 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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 184. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract, education of the teacher' mother and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XIII: Estimation of the effect of teachers' gender stereotypes

Dependent Variable: Track choice Vocational						
Heterogeneous effects by	**Student Characteristics**				**Interaction time with teacher**	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	0.053	0.048	0.052	0.051	0.037	0.063
	(0.073)	(0.072)	(0.076)	(0.073)	(0.074)	(0.073)
Fem*Teachers' Stereotypes(IAT)	0.024**	0.024	0.051**	0.021**	0.016	0.046**
	(0.010)	(0.016)	(0.023)	(0.010)	(0.012)	(0.022)
Fem*IAT*HighEduM		-0.005				
		(0.019)				
Fem*IAT*Top tercile Math6			-0.055**			
			(0.025)			
Fem*IAT*Middle tercile Math6			-0.019			
			(0.028)			
Fem*IAT*Immigrant				0.021		
				(0.029)		
Fem*IAT*Extended School Day					0.034	
					(0.024)	
Fem*IAT*Same Math Teacher						-0.027
						(0.025)
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.	8451	8451	8451	8451	8451	8451
R^2	0.211	0.212	0.216	0.211	0.211	0.211

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on the choice of vocational high-school track 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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 551. The variable "Fem" indicates the gender of the student, "HighEduM" wether the mother has at least a diploma, "tercile Math6" is the tercile of standardized test score in math in grade 6 and "Immigrant" is a dummy equal to 1 if 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' 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.XIV: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - class FE regression

Dependent Variable: reading standardized test score in grade 8					
	(1)	(2)	(3)	(4)	(5)
Fem	0.196***	0.082***	0.079***	0.087***	0.534**
	(0.021)	(0.016)	(0.016)	(0.032)	(0.265)
Fem*Ita Teacher Stereotypes			0.008	0.005	0.008
			(0.015)	(0.015)	(0.016)
Fem*Ita Teacher Fem					-0.009
					(0.052)
Fem*Ita Teacher Born North					-0.016
					(0.031)
Std Ita grade 6		0.729***	0.729***	0.692***	0.690***
		(0.016)	(0.016)	(0.018)	(0.018)
Constant	-0.007	-0.061***	-0.061***	-0.224***	-0.223***
	(0.011)	(0.008)	(0.008)	(0.024)	(0.024)
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	8622	8622	8622	8622	8622
R^2	0.185	0.592	0.592	0.603	0.604

Notes: This table reports OLS estimates of equation 1, where the dependent variable is Italian standardized test score in grade 8; the unit of observation is student i, in class c taught by Italian teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 355. The number of fixed effects (classes) is 510. 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. All columns include the interaction between reading standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, degree with honour, update courses, age, type of contract, and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XV: 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)
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Fem	-0.205***	-0.056**	-0.071***	-0.040	0.172***	0.059**	0.052**	0.089
	(0.026)	(0.023)	(0.025)	(0.106)	(0.027)	(0.024)	(0.025)	(0.076)
Fem*Ita Teach Stereotypes		0.023	0.022	0.008			0.016	0.005
		(0.022)	(0.022)	(0.023)			(0.023)	(0.024)
Fem*Math Teach Stereotypes			-0.035*	-0.043**		-0.029	-0.028	-0.033
			(0.020)	(0.020)		(0.022)	(0.022)	(0.020)
Std Math grade 6		0.720***	0.720***	0.695***				
		(0.017)	(0.017)	(0.018)				
Std Ita grade 6						0.711***	0.710***	0.675***
						(0.023)	(0.023)	(0.024)
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.	4890	4890	4890	4890	4890	4890	4890	4890
R^2	0.188	0.605	0.606	0.615	0.178	0.584	0.584	0.597

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 Italian and math teachers. Standard errors are robust and clustered at math teacher level in parentheses. 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. All columns except 1 and 5 include the interaction between standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher charactersitics. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Math Grade end middle school**				**High Math Grade (>than6)**			
Fem	0.204***	0.348***	0.336***	0.228	0.069***	0.119***	0.116***	0.019
	(0.027)	(0.026)	(0.051)	(0.191)	(0.011)	(0.011)	(0.023)	(0.081)
Fem*Teacher Stereotypes		-0.016	-0.018	-0.033		-0.004	-0.004	-0.013
		(0.027)	(0.028)	(0.028)		(0.011)	(0.011)	(0.011)
Fem*Math Teacher Fem				-0.014				-0.029
				(0.066)				(0.028)
Fem*North Math Teacher				0.019				-0.000
				(0.051)				(0.023)
Fem*Advanced STEM Teacher				-0.029				0.011
				(0.060)				(0.027)
Std Math grade 6		0.759***	0.717***	0.718***		0.262***	0.244***	0.244***
		(0.018)	(0.019)	(0.019)		(0.006)	(0.007)	(0.007)
Constant	6.977***	7.285***	7.276***	2.693***	0.508***	0.542***	0.547***	-0.626***
	(0.013)	(0.013)	(0.043)	(0.255)	(0.006)	(0.005)	(0.019)	(0.112)
Obs.	9073	9073	9073	9073	9073	9073	9073	9073
R^2	0.118	0.435	0.448	0.449	0.103	0.331	0.345	0.347
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	No	Yes	Yes
Math 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 are robust and clustered at math teacher level in parentheses; the number of clusters is 303. The number of fixed effects (classes) is 530. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVII: Estimation of the effect of teachers' gender stereotypes on retention rate and on the probability of doing the standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable	**Retention Rate**				**Doing Test in Grade 8**			
Fem	-0.027***	-0.035***	-0.046***	-0.039	0.014***	0.022***	0.007	0.005
	(0.005)	(0.005)	(0.011)	(0.034)	(0.005)	(0.006)	(0.007)	(0.034)
Fem*IATMath		0.004	0.004	0.003		0.002	0.006	0.007
		(0.005)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)
Std Math grade 6		-0.053***	-0.043***	-0.043***		0.040***	0.024***	0.024***
		(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
Constant	0.060***	-0.305***	-0.337***	2.395***	0.939***	0.877***	1.069***	1.043***
	(0.002)	(0.002)	(0.011)	(0.049)	(0.002)	(0.002)	(0.011)	(0.049)
Obs.	9812	9812	9812	9812	9812	9812	9812	9812
R^2	0.100	0.139	0.156	0.158	0.179	0.203	0.388	0.389
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. 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 retention rate in columns 1-4 and the probability of doing the standardized test score in grade 8 in columns 5-8; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 304. The number of fixed effects (classes) is 556. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' mother. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

	Std Math 8th grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.058***	0.040	0.011	-0.078***	-0.006	-0.035
	(0.017)	(0.104)	(0.105)	(0.016)	(0.104)	(0.104)
Fem*No Dif Innate Ability	0.031*	0.026	0.027			
	(0.017)	(0.017)	(0.017)			
Fem*Teacher Stereotypes			-0.040***			-0.045***
			(0.015)			(0.015)
Fem*WVS				0.009	0.002	-0.001
				(0.032)	(0.031)	(0.030)
Std Math grade 6	0.720***	0.694***	0.694***	0.722***	0.697***	0.697***
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	Yes	Yes	No	Yes	Yes
Obs.	8720	8720	8702	8941	8941	8907
R^2	0.619	0.627	0.627	0.620	0.628	0.628

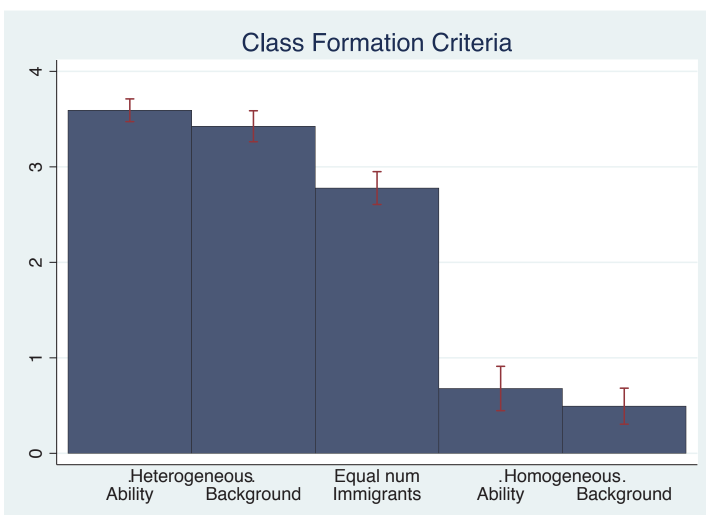
Notes: This table reports OLS estimates, where the dependent variable is the standardized test score in grade 8; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses. The variable "Fem" indicates the gender of the student, "No Dif Innate Ability" assumes value -1 if the teacher think there are no innate differences in math abilities between men and women, 0 if there are few and 1 if there are up to several differences. "WVS" is the World Value Survey answer to the question on the right to access to jobs for men and women. 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. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honour, update courses, age, type of contract and education of the teacher' 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 update courses offered to teachers. We received the questionnaire completed by 82 principals. 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 she has been assigned the direction of the institution in our sample temporarily until new staff is hired. This practice is wide-spread in Italy.

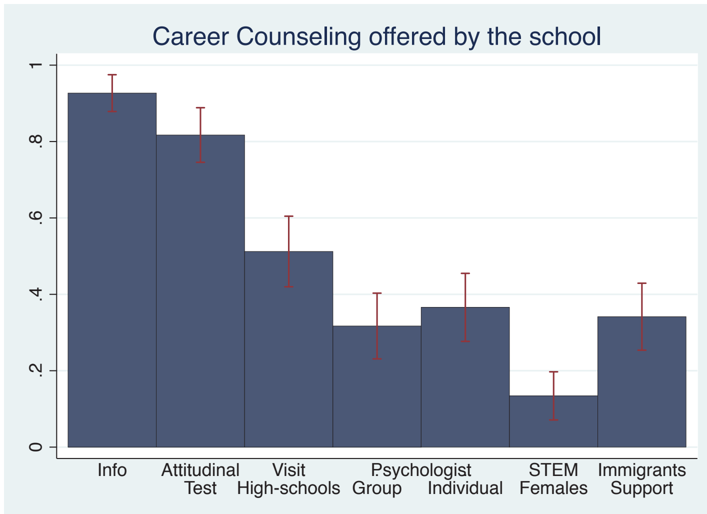
Among 81 principals who completed the questions on class formation, 64 percent consider the heterogeneity across classes in the ability level as “Extremely important” and 33 percent as “Important”. The heterogeneity across classes in the socio-economic status is considered as “Extremely important” by 60 percent of principals and “Important” by 29 percent. The equal allocation of immigrant across classes is considered “Extremely Important” by 25 percent of principals and as “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 with principals



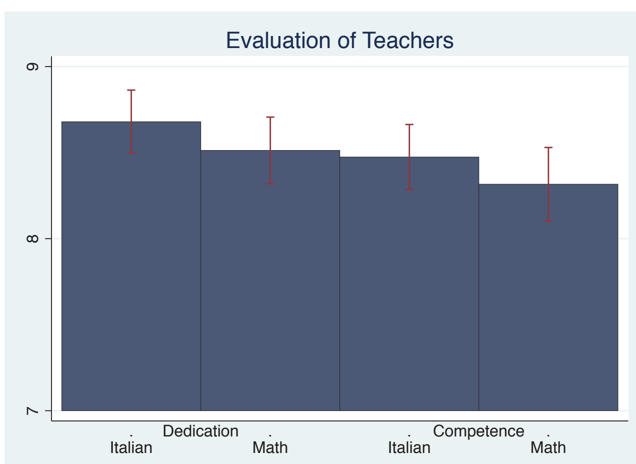
An entire section of the questionnaire is dedicated to career counseling practices done in the school and the share of schools offering the different services is reported in Figure B.II. Most institutions declare to offer information on high-schools curricula and attitudinal tests that help high-school choice. Around one third of schools organize meetings with psychologists at individual or group level to induce students to reflect on this important choice. However, very few institutions try to sensitize 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 Italian teachers in the middle school: generally principals are generous with the evaluation, but if anything, math teachers are considered as marginally less committed to their work and with lower level of competence than their Italian 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

We invite teachers to complete a seven-block IAT that was following the schematic overview presented in Figure C.I. Half of the teachers, randomly selected at 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 categories and stimuli are shown in Figure C.II, while a screenshot of the tablet is shown in Figure C.III. Furthermore, teachers were asked to complete a race IAT with male names and female names of Italian and immigrants. The order of race and gender IATs was randomized at individual level. In Table A.II, 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 that perform the order compatible and incompatible test. Hence, in all regression where there are no class (and therefore teacher) fixed effects I control for the order of IATs.

The blocks used to calculate the IAT score are blocks 3, 4, 6, and 7. The number of words that need to be categorized in blocks 3 and 6 are 20, while in blocks 4 and 7 are 40, as in the standard IAT 7-blocks. The scoring procedure follows the guidelines of the improved scoring algorithm defined by Greenwald et al. (2003).

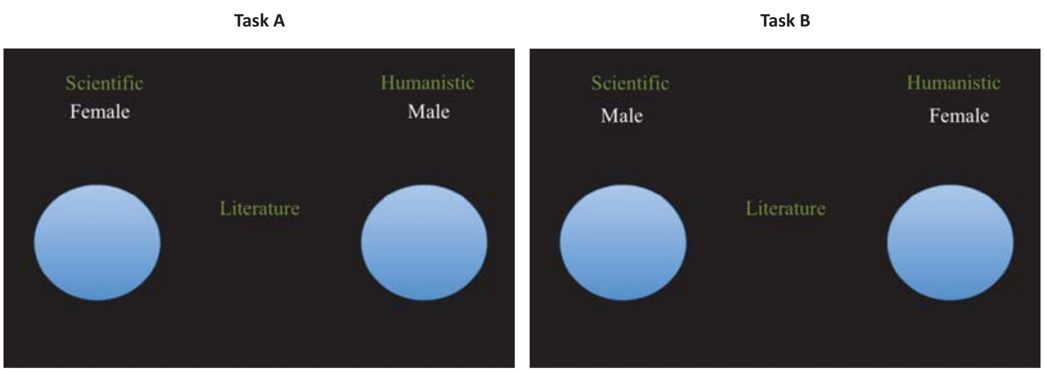
Blocks	Left Categories	Right Categories
1	Maschio (Male)	Femmina (Female)
2	Scientifico (Scientific)	Humanistic (Umanistico)
3	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
4	Maschio (Male) Scientifico (Scientific)	Femmina (Female) Humanistic (Umanistico)
5	Humanistic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

The stimuli presented within each category are summarized in Figure C.II. The improved IAT scoring procedure is used to calculate the bias from the reaction time of individuals to different associations and this method has been shown to be better account for method variance (Greenwald et al., 2003).

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

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

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*"

D Sample Selection

The purpose of this appendix is to investigate two issues related to sample selection: students who did not attend the standardized test score in grade 6 (or their test score was not correctly matched with data from the Ministry of Education) and teachers who did not complete the IAT test either because they were not teaching in the school anymore in 2016-17 or because they did not come to the meeting with enumerators.

In the school involved in this research project, 21,054 students attended the middle school exam in the years between 2013 and 2015. However, we do not have information on their initial test score for 12% of students. Table D.I presents the differences between background characteristics and track choice of students, who attended the standardized test score in grade 6 and students who did not. The latter have an almost double probability of being immigrant and late in the school curricula in terms of grade achieved. They are in general more disadvantaged in terms of family background, achievements and they tend to choose a lower-level high -school track.

Table D.I: Summary Statistics of students by attendance to standardized test in grade 6

	Attendance	No Attendance	Diff.	se
Female	0.480	0.504	-0.024*	(0.011)
Late in school	0.367	0.093	0.274***	(0.007)
Immigrant	0.371	0.185	0.186***	(0.009)
Second Gen. Imm	0.081	0.079	0.002	(0.006)
High Edu Mother	0.331	0.428	-0.098***	(0.011)
High Occ Father	0.130	0.166	-0.036***	(0.008)
Med Occ Father	0.207	0.304	-0.097***	(0.010)
Std Math 8	-0.360	0.057	-0.417***	(0.021)
Std Reading 8	-0.471	0.074	-0.545***	(0.021)
High-school Track: Scientific	0.187	0.253	-0.066***	(0.010)
High-school Track: Classic	0.043	0.059	-0.016**	(0.006)
High-school Track: Other Acc	0.185	0.216	-0.031**	(0.010)
High-school Track: Technical Tech	0.196	0.185	0.011	(0.009)
High-school Track: Technical Eco	0.145	0.141	0.003	(0.008)
High-school Track: Vocational	0.245	0.146	0.099***	(0.009)
Track recommendation: Scientific	0.091	0.140	-0.049***	(0.009)
Track recommendation: Vocational	0.519	0.340	0.179***	(0.012)
Born South	0.046	0.029	0.017***	(0.004)
Observations	18586	2468		

Administrative data from MIUR and INVALSI.

The second issue regards data availability on teacher stereotypes. Indeed, enumerators were able to collect information only on around 80% of teachers currently working in the schools. Furthermore, teachers teaching to the three cohort graduating between 2013 and 2015 may have moved to a different school. Hence, we are able to obtain complete student-teacher data only on 9,309 students, 50% initial 18,586 students. Table D.II provides evidence of whether student characteristics are correlated with the attendance of teachers of the IAT survey, within school. We find no significant difference among students of the school whose teacher was present and not present at IAT survey in terms of math ability, immigrant status and family background. There is only a small, but statistically significant, probability of male students being associated with a professor who completed IAT survey. Finally, in column 1, we also correlate the probability of teachers completing IAT with the type of class in terms of school hours without finding statistically significant effects.

Table D.II: Correlation between teacher attendance of the survey and student characteristics

Dependent Variable: Teacher who completed IAT								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Long School	0.055							0.054
	(0.043)							(0.043)
Female		-0.014**						-0.013**
		(0.006)						(0.006)
Std Math 6			0.004					0.001
			(0.005)					(0.008)
Std Math 8				0.006				0.005
				(0.005)				(0.008)
Immigrant					0.012			0.009
					(0.010)			(0.010)
HighEduMother						0.006		0.009
						(0.012)		(0.012)
HighOccFather							-0.015	-0.020
							(0.014)	(0.014)
MedOccFather							-0.000	-0.003
							(0.009)	(0.010)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18530	18586	18586	18586	18586	18586	18586	18530
R^2	0.173	0.171	0.171	0.171	0.171	0.171	0.171	0.173

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports OLS estimate, where the dependent variable is a dummy variable which assumes value 1 if the teacher completed the IAT; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at class level in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively. For few observations, we do not know the number of hours at school. This is the reason why the number of observation is slightly different in columns 1 and 8.

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 is 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 stereotype. 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 is a weakly decreasing function of teachers' stereotype, 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

$$u_i = \theta_i k(\alpha_i, e_i) - c(e_i) \tag{1}$$

where u is differentiable and the sufficient conditions for a local interior maximum hold, k is the benefits function, which depends on ability (α_i) and effort (e_i), and the cost c is paid according with the level of effort e exerted by individual i. The component θ_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:

$$e_s^* = \frac{\theta k_{e\alpha} \alpha_s}{-(\theta k_{ee} - c_{ee})} \gtreqless 0 \tag{2}$$

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[1], 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 teachers' stereotypes. However, if $k_{e\alpha} < 0$, then $e_s^* > 0$. This theoretical results may explain why girls in the top of the initial ability distribution have slightly higher, but yet 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 (α_i) and teacher investment toward pupils, in the form of either time or encouragement, (β_{t_i}). The latter is an additional channel through which teacher bias (s_t) may impact on students' performance and choices. I define β_{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.

[1]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:

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

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:

$$e_s^* = \frac{\theta(k_{e\alpha}\alpha_s + k_{e\beta}\beta_s)}{-(\theta k_{ee} - c_{ee})} \gtreqless 0 \tag{4}$$

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

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.

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.

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