Stereotypes and Self-Stereotypes: Evidence from Teachers' Gender Bias*

Michela Carlana†

October 2017

JOB MARKET PAPER

LATEST VERSION HERE AND APPENDIX HERE

Abstract

I study the impact of exposure to gender-biased teachers on student achievement and self-confidence. The gender gap in math performance substantially increases when students are quasi-randomly assigned to teachers with stronger stereotypes (as measured by an implicit association test). The effect is driven by lower performance of female students, while there is no impact on males. Teacher bias induces females to self-select into less demanding high-schools, following the track recommendation of their teachers. Finally, teacher bias has a substantial negative impact on females' assessment of their own math ability. These findings are consistent with a model whereby ability-stigmatized groups underperform and fail to achieve their potential.

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

*I am grateful to Alberto Alesina, Eliana La Ferrara, Nicola Gennaioli and Paolo Pinotti for insightful comments and encouragement. I thank for their useful suggestions Ingvild Almas, Thomas Le Barbanchon, Pamela Giustinelli, Selim Gulesci, Giampaolo Lecce, Andreas Madestam, Valerio Nispi Landi, Laura Ogliari, Jonathan de Quidt, Dan-Olof Rooth, David Stromberg, Jackob Svensson, Guido Tabellini, Marco Tabellini, Anna Tompsett, Diego Ubfal and seminar participants of Oxford Development Workshop 2017, 32nd AIEL Conference, SSE Human Capital Workshop 2017, IIES Brownbag and Bocconi F4T Brownbag. 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.

†PhD Candidate, Department of Economics, Bocconi University (e-mail: michela.carlana@unibocconi.it).

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). Several studies have shown that math test scores are good predictors of future occupation and earnings (Altonji and Blank, 1999). Gaining a better understanding of the reasons behind the emergence of the gap in math skills is of first-order importance to explain the enduring gender differences in readiness for science, technology, engineering, and math (STEM) universities and the underrepresentation of women in these highly profitable fields (Card and Payne, 2017).

The gender gap in math performance is generally attributed to either biologically based explanations in brain functioning or social conditioning.[1]. In this paper, I focus on the latter and I study whether exposure to gender stereotypes of teachers during middle school can affect math achievement, track choice, and self-confidence of boys and girls. According to social psychology literature, teachers believe math is more difficult for girls than equally achieving boys (Riegle-Crumb and Humphries, 2012; Tiedemann, 2002). Gender stereotypical beliefs are pervasive and deeply-held in most societies: women are believed to be worse than men in mathematics and arithmetic, even in tasks in which both genders perform equally well on average (Bordalo et al., 2016; Reuben et al., 2014).[2] However, our understanding of the role of gender stereotypes on educational outcomes is limited by the difficulty in measuring stereotypes. Also, no evidence exists on the effect of gender bias on students' self-confidence. This paper addresses both of these gaps.

Analyzing the role of teacher stereotypes on student outcomes presents two main challenges: identification and the measurement of stereotypes. I tackle the former by exploiting quasi-random assignment of students to teachers with different level of bias, within the same school. I measure stereotypes by collecting teacher bias using an Implicit Association Test (IAT). This 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 (Reuben et al., 2014; Glover et al., 2017; Lowes et al., 2015; Burns et al., 2016).

I find that the effect of teachers' gender stereotypes is negative and quantitativly signif-

[1] For instance, Baron-Cohen (2003) elaborates the "empathizing-systemizing theory" whereby there are evolutionary differences among genders: females are stronger empathizers and males are stronger systemizers.

[2] Stereotypes are overgeneralized and simplified representation of differences between groups, which may hold a *kernel-of-truth* (Bordalo et al., 2017). For instance, the belief that *women are worse than men in math* is based on the empirical evidence that girls lag behind in math test-scores in most countries by the age of 14.

icant. First, I show that the gender gap in math performance during middle school increases by 34 percent when students are assigned to teachers with one standard deviation higher bias. The gender gap in math improvements almost triples in classes where the math teacher has a “pro boys” attitude compared to classes in which she or he has a “pro girls” attitude.[3] The effect is driven by lower performance of females when assigned to biased teachers, while males are not affected by exposure to gender bias. Those lagging behind the most when assigned to biased teachers are girls from disadvantaged backgrounds and with lower initial level of achievement. Second, I provide evidence that teacher bias is correlated with their high-school recommendation to pupils and it induces females to undertake less demanding tracks. Finally, I discuss two mechanisms behind the negative impact of teacher bias on student achievements: self-stereotypes and pupil-teacher interaction. I show that teacher stereotypes have a substantial negative impact on girls’ self-confidence in math. The findings are consistent with a model whereby ability-stigmatized groups underperform failing to achieve their potential.

To perform the analysis, I build a unique dataset, combining administrative information on pupils from the Italian Ministry of Education and the National Institute for the Evaluation of the Italian Education System (INVALSI) with a newly collected questionnaire on students and teachers in Italy. I survey more than 1,400 math and literature teachers, working in 103 schools in the North of Italy. As measure of gender bias, I collect Gender-Science IAT. 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 brain (Lane et al., 2007). Implicit bias has been found to correlate with many outcomes in the real world and in laboratory experiments, related for instance to hiring decisions (Reuben et al., 2014; Rooth, 2010). In addition to IAT scores, I have collected detailed information on teacher characteristics, such as family background, teaching experience and explicit gender beliefs. These data are matched with student performance in math and reading standardized test scores, family background, high-school track choice and teachers’ track recommendation. Finally, the dataset is complemented by original information on self-confidence for a sub-sample of students.

I present evidence from two empirical strategies. The first one investigates the impact of teacher bias on the gender gap within the class. I include class fixed effects, which absorb all characteristics of peers, school environment, and teachers, including the level of gender

[3] I consider the thresholds defined by Greenwald et al. (2003) to identify teachers with “pro boys”, “pro girls”, “without bias” attitude (the latter is a IAT score between -0.15 and 0.15). The gender gap in math performance is -0.035 standard deviations in “pro girls” classes and -0.10 standard deviations in “pro boys” classes. The increase by 34 percent in the gender gap when students are assigned to teachers with one standard deviation higher bias corresponds to an increase of 0.03 standard deviations with respect to an average gap in test scores generated during middle school of 0.08 standard deviations.

stereotypes. I exploit variations in performance and track choice between boys and girls enrolled in the same class.[4] In the second empirical strategy, I compare students of the same gender, enrolled in the same school and cohort, but assigned to teachers with different level of bias. Both identification strategies rely on the quasi-random assignment of students to teachers with different level of bias. I provide supporting evidence showing that baseline characteristics of students are not systematically correlated with teacher bias.

This paper makes three contributions. First, I show that implicit bias correlates with several "expected" observable characteristics, as gender, field of study and cultural stereotypes in the place of birth of individuals (as measured by the World Value Survey and female labor force participation).[5] IAT does not correlate with variables such as gender of own children, teacher quality and experience.[6] Second, the paper provides evidence on the relevance of social conditioning in affecting the gender gap in math achievement and high-school track choice. More precisely, it uncovers the role of implicit bias in the context of education economics and pupil-teacher interactions. Third, it shows the influence of teachers on self-stereotypes and self-assessment of own math ability. This is a crucial channel to explain the underperformance of girls in math when assigned to more biased teachers.

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 (Bertrand et al., 2005; Nosek et al., 2002). In particular, we may expect that teachers do not explicitly endorse gender stereotypes, but their implicit bias, embedded in their own experiences since childhood, affects their interaction with pupils. I collect IAT scores and I further examine the determinants captured by this test and the reaction time to stimuli.

Some interesting cross-countries evidence shows a correlation between gender gap in mathematics and gender equality. Guiso et al. (2008) and Nosek et al. (2009) provide evidence that gender gap in math performance is wider in those countries with low women empowerment and higher implicit gender bias measured by IAT, respectively.[7] The economics literature ana-

[4] 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.

[5] 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".

[6] This has important implications for the estimation: when teachers controls or class fixed effects are not added in the regression, we need to consider teacher bias as including also characteristics correlated with IAT scores.

[7] Guiso et al. (2008) use four measures of gender equality: World Economic Forum's Gender Gap Index (GGI), World Values Surveys (WVSs), labor force participation of women and women's political participation measured by World Economic Forum. They find consistent results with all these measures.

lyzing the impact of gender stereotypes of teachers on student outcomes has mainly focused on either self-reported measures (Alan et al., 2017) or bias in grading, i.e. the gender differences in grades given in blind vs. open evaluations (Lavy and Sand, 2015; Terrier, 2015).[8] 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, the measure of teacher bias is created without relying on data on student performance, which may capture random variation in unobservable characteristics of boys and girls, potentially correlated with future outcomes of pupils.

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, scientific educational and occupational choices (Kugler et al., 2017; Reuben et al., 2015; Coffman, 2014).[9] 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 suffer reduced 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 analyzed, providing information on the Italian institutional background. Section 3 describes the data available on both students and teachers. Section 4 presents the estimation and identification challenges. The main results of the paper are presented in Section 5 and mechanisms are discussed in Section 6. Finally, Section 7 concludes. All supplementary material is provided in the Appendices.

2 Setting

In the Italian educational system, middle school lasts three years from grade 6 to 8. Students are assigned to classes at the beginning of grade 6 and they stay with the same peers for three years.[10] 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 insti-

[8] 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.

[9] Niederle and Vesterlund (2010) point out that gender differences in competitiveness may have some distortionary effects and exaggerate the advantage of males in math, especially in the right tail of the distribution of test scores.

[10] There are only few exceptions: students may be transfer to a different school or be required to repeat a grade. This affects less than 5% of students.

tution.[11] The general criteria mentioned directly 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). I also collect additional information directly from the principal on how classes are formed, which is described in details in Appendix B. School principals report that the most important aspect in the class formation process is the comparability across classes and heterogeneity within class in the same school.[12] 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 they are all paid equally in a centralized system. Teachers' allocation across school is determined by seniority: when they accumulate years of experience, they tend to move close to their home town and away from disadvantaged backgrounds (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.[13] Students receive grades by teachers at the end of each semester, which may be affected not only by performance, but also by other factors as diligence, effort and improvements over time. Grades are given in a scale up to 10, where the pass grade is 6.

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).[14] 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 it will affect 1/6 of the final score of students at the end of middle school. However, this final grade has no relevant impact for the enrollment in high-school or for the future educational career of students.

After middle school, students self-select into three different tracks: academic oriented

[11] 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").

[12] An analysis of Ferrer-Esteban (2011) shows that there is ability grouping across classes within schools occurs almost exclusively in the South of Italy, while all schools in my sample are from the North.

[13] 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

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

("liceo"), technical and vocational high-school. Each type of school is divided in several sub-tracks: 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. Giustinelli (2016) has shown that child's enjoyment of the curriculum is one of the most important determinants of high school choice. 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.[15]

3 Data

During September 2016, I invited 156 middle schools to take part in a research project regarding "The role of teachers in high-school track choice," out of which 91 accepted and provided all information necessary for my study. 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.[16]

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 directly detailed information on teachers, including *implicit bias* measured by the Gender-Science Implicit Association Test (IAT), and on students' self-assessment of own ability in different subjects. Administrative

[15] Author's calculation on MIUR data.

[16] More precisely, in 103 schools we obtain 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). In 2 cases, the principal explicitly stated they did not want to give access to INVALSI data. In most of the cases, the authorization (with all correct data) was not sent in time for the extraction of data from INVALSI. Finally, the number of schools according with 2011 data were 145. However, some of them where divided in different institutions and we follow all of 156 of them over time.

data from MIUR contained information on gender, place of birth, high-school track choice, and their track recommendation to students. INVALSI provides information on standardized test scores and family background.

3.1 Teachers: Gender Stereotypes and Other Characteristics

From October 2016 to March 2017, I conducted a survey of around 1.400 math and Italian teachers. The questionnaire was administered directly one-to-one by enumerators using tablets in a meeting held in school buildings, in most of the cases in the early afternoon. 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.[17] There was no direct reference to gender bias. The time to complete the survey was around 30 minutes and participating teachers did not receive compensation. Among all math and Italian teachers working in the schools involved in this research, around 80 percent completed our survey.[18] The survey is divided into two parts: the Implicit Association Test (IAT) and a questionnaire.

Gender-Science Implicit Association Test

In this research, the main focus is on *implicit* gender bias, using a measurement tool developed by social psychology called Implicit Association Test (IAT) (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.[19] 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 (specifically, gender and scientific/humanistic subjects). Enumerators administered the test using touch screen tablets and they interact directly one-to-one with teachers. Subjects were presented with two sets of stimuli. The first set of stimuli were typical Italian names of females (e.g. Anna) and males (e.g. Luca), and the second set were subjects related to scientific fields (e.g., Calculus) and humanistic fields (e.g., Literature). One word at a time appears at the center of the screen and

[17] The data collection was conducted for a broad research project involving also an ongoing work in which we study teachers' racial bias (Alesina et al., 2017).

[18] 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 scope of the Implicit Association Test.

[19] 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.

individuals are instructed to categorize them to the left or the right according with different labels displayed on the top of the screen (for instance on the right the label "Females" and on the left the label "Males"). Subjects are required to categorize the words as quickly as possible for seven blocks, i.e. seven rounds. To calculate the score, two types of blocks are used: in the first type, 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" blocks), while in the second type of blocks, 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" blocks).[20] The order of the two types of blocks is randomly selected at individual level. Appendix Table A.1 presents the correlation between IAT score and whether the first task was order compatible or incompatible. The effect is small in magnitude and it disappears when controlling for school fixed effects.[21]

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 gender IAT) predict employers' bias 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). Also in the context of race implicit bias, studies have shown the relevance of IAT scores in call-back rates of minority job applicants (Rooth, 2010) and in affecting job performance of minorities (Glover et al., 2017).

There is a lively debate in the literature on how to interpret IAT scores and to what extent they are capturing stable characteristics that do not vary over time (Banaji et al., 2004; Greenwald et al., 2009).[22] Thanks to a broad set of individual level information on teachers, I will contribute to this debate by analyzing the correlation between observables and IAT score in

[20]The number of words that appear in the two types of evaluation blocks are 120. As in the standard IAT with a seven-block structure, individuals are asked to categorize only female and male words in the first block, only scientific and humanistic subjects in the second and fift, while blocks three/ four and six/seven are those described in details and used for evaluation. Detailed explanation is provided in Appendix C.

[21]Each teacher perform both gender and race IAT. The order was randomized at individual level. In the same Appendix Table, I control whether the order of IATs has an effect on the score. The correlation is low and indistinguishable from zero. However, in all regressions I will control for ordering factors (even if they have no impact on the estimates).

[22]In particular, it has been shown that race bias (as measured by IAT) decreases after subjects viewed pictures of admired African Americans and disliked White Americans (Dasgupta and Greenwald, 2001).

Section 4.2.

One of the critiques of implicit bias is that people may have a specific context in mind when completing the IAT, which may differ from the context the researcher wants to analyze. If anything, it would increase the noise of the measurement inducing an attenuation bias. However, in my case, this is not a relevant issue. The measure of bias I collect is strongly related to the schooling context and teachers are interviewed directly inside the school building.[23] Furthermore, individuals complete the survey in the presence of an enumerator and therefore I am sure of the identity of who completed the survey.[24]

Teachers' Questionnaire

After the Implicit Association Tests, enumerators invited teachers to complete a questionnaire with 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 forms of 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"*.[25] Participants are in general reluctant to explicitly endorse gender stereotypes about differences in innate ability and employment (Nosek et al., 2002) due to social desirability bias in the responses. These aspects are potentially 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 confirmed all this information using data provided directly by schools and their websites.

[23] An example in which this may be an issue is the following. Assume I was interested in evaluating the bias toward obese people in the work environment and I collected IAT associating "obese people" and "thin people" with "good" vs. "bad". The positive attitude captured by IAT of a person toward obese people may be due to the fact that his/her mother is obese and he/she loves her. In the job environment, however, the same person may have a neutral attitude toward obese people. This would induce a bias in our measure of attitude toward obese people in the workplace. The context the person has in mind when completing the IAT may have an important effect on the result. In our case, the context of IAT is the same as the outcome I want to evaluate.

[24] A less-expensive and time-consuming alternative could have been sending the survey by email. However, the potential drawbacks were low response rate and uncertain identity of the individual completing the survey.

[25] I also collected information about potential factors that may influence females' scientific track choice (interest for STEM, ability in math, low self-esteem, parents' influence toward different tracks, cultural stereotypes) using a scale of 1 to 5. I also elicited the average performance of their students in the standardized test score, by gender, including a question in which teachers were asked to reveal how sure they felt about their answer.

Descriptive Statistics on Math Teachers

The dataset includes 537 math teachers, 855 Italian teachers and 31 teachers of other subjects. The main focus of this paper is on the impact of math teachers gender stereotypes on the performance in the subject they teach. Among these 537 teachers, we restrict the main analysis to 301 teachers ("matched sample") that were hired by the same school in the school year 2014-15 and for which we have student data [26]. Appendix Table A.2 shows the balance table of the differences between the sample of teachers matched (301 teachers) and the other 236 math teachers who completed the IAT. As expected, teachers not matched are around 9 years younger, 40 percent less likely to have full-time contract and they have 12 years less of experience in teaching. However, as it can be clearly seen also from Appendix Figure A.1, 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: 0.946).

Table 1 reports descriptive statistics on math teachers. 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 (65%) of math teachers are born in a city in the North of Italy where the study took place, 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: 24% studied math, physics and engineering. At the bottom of Table 1, I report the summary statistics of explicit bias questions described in details in Appendix C. The variation in the answers on the equality of access to labor market of men and women and about innate gender difference is ability is low, potentially also due to social desirability bias: for instance, less than 2% of the interviewed teachers respond that they agree with the statement that women have less right to jobs than men when opportunities are low. It may be difficult to obtain revealed bias, given the widespread explicit rejection of stereotypes and a related reluctance of participants in revealing their bias, especially if interviewed as "teachers" in the presence of enumerators.

Based on IAT scores, math teachers are slightly gender biased: indeed, a positive IAT score indicates a stronger association between males with scientific subjects and female with humanistic subjects. For ease of interpretation of our results, I standardize the IAT score to have mean zero and variance one throughout the paper. Considering the thresholds typically used in the social psychological literature, 25% of teachers are slightly or moderately in favor of girls, 30% present little to no bias, 19% show slight bias against girls and 26% show moderate to

[26]As specified in section 3, for 12 schools we did not obtain this authorization on time or there was a mistake in the authorization form. Furthermore, we lost some observations because some schools changed the official code (called "meccanografico") over the years of our sample and INVALSI guarantees access to data only for school codes whose principal has signed the authorization.

severe bias against girls. [27] The sample of 1164 Italians used by Nosek et al. (2009) that decided to take the IAT online in a similar Gender-Science test have an average score of 0.40 (SD 0.40): the score of math teachers is on average lower than this sample (mean 0.09, SD 0.37, as shown in Table 1), while Italian teachers are very close to it (mean 0.39, SD 0.39, as shown in Table E.1). [28] Interestingly, the great majority of math teachers are women and this may have important implications for the association of scientific subjects with gender.

3.2 Students: Self-Stereotypes and Administrative Data

I use individual level information from the Italian Ministry of Education and from INVALSI for three cohort of students enrolled in grade 6 between school year 2010-11 and 2012-13. [29] 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, around two months before the end of middle schools, are asked to complete a survey about their track choice. In particular, they need to mention all subjects they will learn during high-school and to report their belief about their own ability in each subject. The potential choices to that answer were: "good", "mediocre", "scarce".[30]

Table 2 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 6. This is the sample that will be used in the empirical analysis of this paper. Appendix D describes in details the sample selection and potential attrition issues. In our 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 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 the average national choices in those years: females are almost 10 percentage points less likely

[27] Greenwald et al. (2003) suggests that a raw IAT score below -0.15 show bias in favor 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.

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

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

[30] These information were collected by Carlana et al. (2017) in a random sample of 47 control schools to evaluate soft skills of students. The specific question exploit in this paper was not used in the paper by Carlana et al. (2017).

to choose an academic scientific track and almost 25 percentage points less likely to enroll in a technical technological track. Girls are more likely to choose an academic track than boys, but not a top-tier academic track (which include classical and scientific tracks). Indeed, one third of females choose a social, linguistic or artistic academic 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 16% of males and 11% of females.[31] Finally, 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 compared to girls.

4 Empirical Strategy

In this section, I present the estimating equation (Section 4.1) and the threat to identification. In Section 4.2, I analyze more deeply teacher characteristics correlated with IAT score that will be important for the estimation as well. Then, I focus in particular on providing evidence in support of exogeneity assumption in the assignment of teachers with different level of bias to students (Section 4.3) and the potential reverse causality issue (Section 4.4).

4.1 Estimating Equation

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

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

where y_{ict} is an outcome of student i in class c taught by teacher t in grade 8 (math standardized test score, track choice, and self-confidence). $Female_i$ is a dummy variable which assumes value 1 if the student i is a girl and $bias_{ct}$ is the standardized value of the gender implicit bias of teacher t assigned to class c in grade 8.[32] I include fixed effects at class level η_c, which absorb the average effect of the bias in class c. Furthermore, for robustness, I include

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

[32] 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.

student characteristics $\mathbf{X}_i$ (parents education and occupation, immigration status and generation of immigration), and teacher characteristics $\mathbf{Z}_{tc}$ (as gender, place of birth, age, teacher quality, type of contract, type of degree achieved and self-reported gender bias) interacted with the gender of student i. Standard errors are 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 bias 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 changes when assigned to teachers with one standard deviation higher bias.[33] 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).[34] 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 $\mathbf{Z}_{tc}$. This is potentially important to partial out differential impact by gender of sex, background, and experiences of teachers. Furthermore, this allows to establish whether the impact of teacher stereotypes on gender gap among classmates can be explained (or attenuated) by teachers' observables, as clarified in Section 4.2.

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 bias level. I investigate whether the impact of teacher stereotypes on gender gap is due to higher performance of boys, lower performance of girls or a combination of both. I estimate the following equation:

$$\begin{aligned} y_{ictsp} = \beta_0 + \beta_1(Female_i \times bias_{ct}) + \beta_2 Female_i + \beta_3 bias_{ct} + \eta_{sp} + \\ + \mathbf{X}_i\rho_1 + (Female_i \times \mathbf{X}_i)\rho_2 + \mathbf{Z}_{ct}\rho_3 + (Female_i \times \mathbf{Z}_{ct})\rho_4 + \varepsilon_{ict} \end{aligned} \tag{2}$$

where η_{sp} are school by cohort fixed effects and standard errors are clustered at teacher level. All other variables are defined as in equation (1).

Institution level characteristics are absorbed by school by cohort fixed effects. The advantage with respect to specification (1) is that we can analyze the impact of teacher stereotypes separately on male students (β_3) and on female students ($\beta_1 + \beta_3$). The drawback is that we cannot control for unobservable characteristics at the teacher or class level: this specification

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

[34] 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.

exploits variation in the level of teacher bias to which students of the same gender in the same school and cohort are exposed.

4.2 Correlation between implicit bias and individual characteristics

In this Section, I present evidence on the correlation between observable characteristics of teachers and IAT scores. This has important implications for the estimation. When teachers controls are not included, we need to consider teacher stereotype as including also characteristics correlated with teacher bias. This point will be carefully stressed analyzing the results.

Figure 1 plots the entire distribution of implicit bias for math and Italian teachers by gender: interestingly, individuals teaching a subject which is stereotypically associated with their gender (i.e. males teaching math and females teaching Italian) are more gender biased. Teachers are more likely to associate own gender with the subject they teach. This result is coherent with findings of Rudman et al. (2001) according to which individuals possess implicit gender stereotypes in self-favorable form because of the tendency to associate self with desirable traits.

The richness of the data collected allows me to associate individual level characteristics of teachers with the results from the Implicit Association Test (IAT) in order to dig deeper into the determinants captured by reaction time to stimuli. Table 3 shows the correlation between math teacher IAT score and their characteristics. Women teaching math are significantly less biased in associating gender with STEM and this explains a substantial portion of the low average IAT score for math compared to Italian teacher. In columns 2-5 (Panel A), I show the association with age, education of teachers' mother, and whether teachers have children. Among this group of comparable adults, implicit stereotypes is not affected by age. Teachers with mothers that graduated from high-school seems to be slightly less biased, even if the effect is imprecisely estimated. Finally, having children, and in particular daughters, do not significantly impact on gender stereotypes.[35]

Gender stereotypical beliefs are rooted in cultural traits, transmitted from generation to generation (Guiso et al., 2006). Indeed, I find that exposure to cultural norms is strongly associated with the IAT score. In column 1 of Table 3 (Panel B), I correlate the implicit bias 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) using World Value Survey data at Italian provincial level.[36] I further investigate how implicit

[35] I also control if the Gender-Science IAT score is correlated with the race IAT score. In the same regression as in Table 3, 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 test does not seem to capture a general "ability" in doing the test.

[36] Italy is a country with low labor market participation of women, but substantial geographic variation across

association are correlated with individual level beliefs and cultural norms in the place of birth. As shown in column 2 of Panel B, participation of women in the labor force in the province of origin of teachers is correlated with the IAT score.[37] In column 3 and 4 of Panel B, I adopt a standard definition of culture based on individual preferences, as measured by the question on the relative rights of men and women to paid jobs when the latter are scarce. I find that the answer in the province in which teachers are born is significantly correlated with their implicit bias. The correlation is low and indistinguishable from zero when considering the answers to the same question given by teachers during the survey I administered. We may suspect that there is a social desirability bias in self-reported measured when professors are interviewed in the school. In column 5 of Panel B, I correlate implicit and beliefs about innate differences in ability between men and women and I find a weak positive correlation (not statistically significant). 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 honor), finding negative point estimates despite high standard errors. Another rough proxy of potential quality of teachers is related to having the tenure (which is associated with higher experience in teaching), and being the professor in charge of math Olympiads in the school.[38] Also in these cases, point estimates are small and indistinguishable from zero.

Appendix Table A.3 shows jointly all correlation presented in separate regressions in Table 3. 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 regarding the absence of a systematic correlation between gender bias of teachers and student characteristics and the absence of systematic grouping of students by socio-economic background and initial ability.

If parents are able to guess who is the teacher with higher stereotyping behaviour, they may try to place their daughter in a different class. Although this seems unlikely, it is also possible that they try to select teachers according to observables which are correlated with IAT score,

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.

[37] The correlation between labor force participation of women the geography is indeed extremely strong in Italy.

[38] In each school, usually only one professor is in charge of math Olympiad and anecdotally she is highly motivated and passionate teacher. Indeed, as shown in Appendix Table A.4, teachers in charge of math Olympics induce higher improvements in test scores of their students.

as gender, place of birth of teachers, years of experience and tenure.[39] In Table 4, I provide evidence that student characteristics are not systematically correlated with the implicit bias of teachers. I would not be able to obtain causal estimates if teachers with higher gender bias are systematically more/less likely to be assigned to females or to females with specific characteristics in terms of parents' education and occupation, place of birth and ability. I might expect that if parents had control over assignment of their children to teachers, daughters of highly educated mothers would have been less likely to be assigned to more biased teachers, within school. Instead, I see that the difference is not statistically significant and the point estimate goes in the opposite direction. In columns 3, 4, 5 and 6, I analyze the correlation respectively with father occupation, immigration background and for the proxy of ability using standardized test scores in reading in grade 6 and I do not find statistically significant correlation. Furthermore, the point estimates are small in magnitude as well. Finally, in the last column, I also include the standardized test score in math in grade 5, before entering middle school, despite the sample size is substantially reduced for data availability issues.[40] The assumption of quasi-random assignment of students in the sample to teachers with different level of gender bias, as measured by the Implicit Association Test, within a school, seems to be supported in the context under analysis.

The result is identical when observations are collapsed at teacher level, as shown in Appendix Table A.5. I also verify that teachers with higher bias are not systematically associated with fewer females in the top of the distribution. I find that this is not the case and, if anything, the sign of the correlation goes in the opposite direction. The results considering the share of female students in the top 10, 20 and 40 percent of the distribution in the standardized test score in grade 6 are shown in Appendix Table A.6.

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 considered as one of the best math teacher. It seems reasonable since, as shown in Appendix Table A.4 for the sample under analysis, his or her students improve the most their math performance in terms of value added, especially females. However, if anything, teachers in charge of math olympics have slightly more gender biased than others, as measured by IAT score (Table 3).

[39] Anecdotally, parents dislike being assigned to a teacher with a temporary contract that may change during the middle school years and has little experience. 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. Almost all teachers included in my analysis have tenure. Hence, among these teachers, the selection on experience is unlikely.

[40] I required standardized test score in math in grade 5. Unfortunately, for reasons related to confidentiality, I have obtained them 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.

Finally, principals must assigned all teachers to a class since they have an exact number of teacher. Hence, they cannot avoid assigning a teacher to a class, even if he or she can guess who is the teacher with higher stereotypes.[41]

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. This objective is spelled out in the official documents on the school websites of most schools and also emerges from self-reported information from principals discussed in Appendix B. It is important to stress that middle school teachers do not teach in elementary school as well. 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. I check whether class assignments are statistically independent with a series of Pearson Chi-Square tests (Lavy and Sand, 2015). 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%. Hence, there is no evidence of systematic non-random formation of classrooms with respect to students' characteristics.

4.4 Reverse Causality

The measure of teacher gender stereotypes was collected between October 2016 and March 2017. Teacher data are matched with students who graduated from middle school between June 2013 and June 2015, as clarified in Figure 2. Similarly to the study of Glover et al. (2017), teacher bias is collected after students in the sample graduated from middle school and therefore after outcomes are realized. The main potential concern is that IAT scores are affected by exposure to students. Indeed, the IAT is expected to be the combination of two aspects: the former is a trait stable over time capturing the influence of cultural norms and experience, while the latter is 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.[42] Our potential

[41] Principals do not have more math teachers available than classes in the school. Since each teacher with a full-time contract teaches three classes, teachers can be assigned to more than one school to cover all their required hours.

[42] 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

concern given the timeline of the analysis is that the three cohort of pupils affect the stable trait of teachers' gender bias.

Reverse causality seems unlikely for several reasons. First, as shown in Section 4.3, teachers with higher bias were not assigned to a differential treatment in terms of student characteristics. I control for student family background, ability in math and reading as measured by standardized test score (see Tables 4 and Appendix Table A.5) and share of females in the top of the math ability distribution (see Appendix Table A.6). Second, under the assumption of monotonic decay of the influence of students to teachers, I would expect a higher effect for the most recent cohort of student. However, results are stable in all three cohorts, as shown in the robustness analysis (Appendix Table A.7). 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. Furthermore, for data availability issues, we do not include in the sample the cohort of student graduating right before the school year in which the test was administered. Each teacher has been exposed on average to 4 classes (around 100 students) after those included in our analysis.[43]

In fact, there is a main advantage from exploiting this timing choice: 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 cohorts of boys and girls in this dataset.

5 The Impact of Teachers' Implicit Bias

In this section, I present the main results of the paper. I focus on the impact of teacher bias on student performance as measured by the standardized test scores in math (Section 5.1) and high school track choice (Section 5.2). Finally, I present some robustness checks and additional outcomes in Section 5.3 before analyzing the mechanisms behind the treatment (self-stereotypes and pupil-teacher interaction).

5.1 Performance in math

By the age of 14, girls are lagging behind in math compared to their male classmates by around 0.22 standard deviations, a result comparable to several other countries (Fryer Jr and Levitt,

in most studies) (Nosek et al., 2007).

[43]Students who were enrolled in middle school in the school year 2015-2016 and 2016-2017 are not included in the sample. Usually, math teachers teach three classes per year (one in grade 6, one in grade 7 and one in grade 8). Hence, teachers are exposed to around 4 different classes and therefore around 100 students after the last cohort of students I analyze and before taking the IAT (i.e. the class in grade 8 in 2015-16 and classes in grade 6, 7, and 8 in 2016-17.

2010; Bharadwaj et al., 2016).[44] As children complete more years of education, the differences between boys and girls gets bigger. 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 5. This paper analyzes what happens to the gender gap when students are quasi-randomly assigned to biased teachers.

Before moving to the causal estimates, Appendix Figure A.3 plots the relationship between teacher bias and math performance of male and female students. Each circle plots the average improvement in math test scores of students assigned to a math teacher with the indicated level of bias, aggregated into bins. The size of the circle indicates the number of observations per bin. These graphs plot the raw data, without removing fixed effect at class or school level. Nonetheless this figure tells a similar story compare to our regression analysis: female students are lagging behind when assigned to math teachers with higher implicit bias.

Table 5 shows the effect of teacher bias 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 bias have 0.027 standard deviations higher gender gap in math performance. Considering an average gap of 0.08 standard deviations, it corresponds to an increase of 34% of the gender difference in performance generated during middle school. 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.

Although the level of teacher bias and all characteristics are absorbed by the class fixed effect, as clarified describing equation (1), column 5 includes the interaction between student gender and teacher characteristics $\mathbf{Z}_{tc}$. If anything, the coefficient of interest *"Fem*Bias Teacher"* 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 teacher bias. 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 children or daughters, whether he or she achieved the degree with *laude*, the type of teaching contract, update courses and appointment as teacher in charge of math Olympics. The latter controls are crude proxies for teacher quality in terms of improvements in standardized test scores, as shown analyzing the relation between value added and teachers observables the Appendix Table A.4.

[44] In Appendix Figure A.2, 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 (0.21 standard deviations as shown in Table 2). Most of the variation in math performance is within classes, coherently with the target in class formation of heterogeneity within class and homogeneity across classes.

As it can be seen in column 5 of Table 5, ceteris paribus, female students assigned to female teachers or to teachers with an advanced STEM degree have slightly lower, albeit insignificantly so, math achievement test scores in grade 8 compared to their classmates. The impact of teacher gender is coherent with the result of Bharadwaj et al. (2016). Having a teacher of own gender helps improve performance, especially at college level (Dee, 2005; Carrell et al., 2010). Finally, teachers born in the North of the country do not have an heterogeneous effect on boys and girls. The results are robust to potential confounding aspects considering all information available on professors from their family background to their professional career.

To give a clearer interpretation, Figure 3 reports the same estimates using a categorical variable instead of the continuous one. 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 and "pro boys"("pro girls") assumes value 1 when implicit bias is higher than 0.15 (lower than -0.15). Being assigned to a teacher with a "pro boys" attitude (45% of teachers) in STEM compared to a teacher with a "pro girls" attitude (24% 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 Appendix Table A.8, considering the thresholds defined by Greenwald et al. (2003) and also whether IAT score is positive or negative. As in Table 5.1, the effect is stronger when controlling for student and teacher characteristics interacted with pupil gender. In columns 4-6, we consider whether IAT score has a positive or negative sign finding similar results.

Are biased teachers worse instructors or are they helping boys to learn math? I next 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 4 shows that having a teacher with strong gender stereotypes has a negative impact on female students, while a bias in favour of girls has a positive impact in their math improvements. The linear approximation presented in this paper seems appropriate. There is no statistically significant impact on male students, throughout the whole distribution of teacher bias. Table 6 mirrors Figure 4: it presents the results of the regression analysis and shows that girls are lagging behind when assigned to more bias teacher, while boys are not affected by teacher stereotypes. The results are robust to the inclusion of the same controls as in Table 5. In this specification the characteristics of teachers are not absorbed by class fixed effects and therefore controls at teacher level, included in columns 5, are particularly relevant. Furthermore, controls for the amount of math hours per week are included in this specification and interacted with the student gender. Indeed, in almost all schools some classes have an extended school day and they spend more time with all teachers, including the math one. Adding all these controls does not significantly impact on the main results.

The differential response by gender is consistent with the previous results in the economic literature: females are negatively affected by teachers of male-typed domains, 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

We now examine which students are most affected by teacher bias. Table 7 shows that the effect of implicit stereotypes is stronger for the most disadvantaged groups of female students, in term of background characteristics. The empirical evidence presented is coherent with stereotype threat model (Steele and Aronson, 1995): individuals with higher risk of conforming to the predicament that *"women are bad at math"* are those more deeply affected. Indeed, male students are not influenced by teacher stereotypes and among females those strongly affected are from disadvantaged backgrounds, especially in terms of initial math achievements.

Based on the estimates in column 2, a standard deviation increase in teacher bias leads to an increase of the gender gap of 0.049 standard deviations among students with low educated mothers and of 0.027 standard deviations among students with higher level of mother education (at least a diploma), although the difference is indistinguishable from zero at usual 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 tercile (-0.070, with standard error 0.027) and turning positive, but not statistically significant, 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 not statistically significant at usual levels.

Why do girls from more disadvantaged backgrounds suffer the most from the interaction with biased teachers? In the case of math-male association, females are more vulnerable to the predicament that *"women are bad at math"* and especially those females with lower initial performance who are at higher risk of confirming the negative expectations on their group. Appendix G presents a conceptual framework that illustrates how teacher stereotypes can differentially affect effort and outcomes of students in the bottom and the top of the ability distribution.[45] 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.

In order to investigate further the second potential explanation, I analyze the heterogeneous effect according to the "quantity" of interaction time between teacher and students. The last two columns of Table 7 analyze whether there are heterogeneous effects in terms of years of exposure and hours per week. Indeed, around 75% of students interact with the math teacher for six hours per week, while the rest for 9 hours per week. 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 lower intensive margins and we 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 2, 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 since buildings are generally separated. Finally, the choice of high-school is strongly correlated with university choice, as discussed in Section 2. From a policy perspective, the scientific academic path is interesting since it 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 teacher bias 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 8, Panel A, shows that girls are 9.4 percentage points less likely than boys to attend a scientific track and equally likely to attend a vocational track. Controlling for the standardized

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

test score in math reduces half of the gap in the choice of scientific track, which is present also in track recommendations received from teachers (Panel B). However, I find a close to zero and insignificant effect of teacher bias on gender gap in scientific track choice (Panel A, columns 2-4) and in the recommendation of teachers toward a scientific track (Panel B, columns 2-4). The inclusion of controls at student and teacher level interacted with the pupil gender do not affect the point estimates of interest. Appendix Table A.9 shows the results estimating equation (2), with school fixed effects instead of class. They confirm the previous evidence of a substantial impact on female students in terms of educational choices.

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 tests, whose performance was not affected by teacher bias, as shown by analyzing the heterogeneous effects in Section 5.1.[46] These female students are likely to have additional academic-oriented role models in addition to their math teacher and a lower vulnerability to the gender stereotypes.

Teacher stereotypes have stronger impact at the bottom of the ability distribution. Indeed, we can observe in columns 6 of Panel A that females, when assigned to a teacher with one standard deviation higher implicit bias, are more likely than their male classmates to attend vocational track by around 2 percentage points. This effect mirrors an analogous differential in teachers' track recommendation toward vocational school as shown by Panel B, columns 6. 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. When exposed to less gender-biased environment, female students are more likely to attend the technical track, instead of vocational (see Appendix Table A.10).

Appendix Table A.11 presents results from the heterogeneous analysis and, as expected, the impact of teacher bias has a stronger effect on the track choice of female students from disadvantaged background. The enrollment of females from the bottom tercile of the distribution increases by 4.3 percentage points for one standard deviation higher bias of the math teacher (which corresponds to a 15.8% increase with respect to the mean value for this group).

5.3 Additional Results and Robustness Checks

Appendix Table A.12 provides evidence of the impact of math teacher bias on reading standardized test scores, presenting the results of estimating equation (1). Although the effect is

[46] 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 is the parental influence (for the summary statistics see Table 1).

significant only including the controls, there are some negative cross-subject spillovers in performance. Additionally, Appendix Table A.13 shows estimates of the impact on math performance of the Italian teacher bias. The gender bias of Italian teacher does not affect the gender gap in math performance. The point estimates are small, indistinguishable from zero and not affected by inclusions of controls either at Italian teacher level or at pupil level.[47] Biased teachers in male-typed domains activate stereotypes on female students. Indeed, gender bias of Italian teachers has no statistically significant impact on reading and math performance of students, neither boys nor girls.

All results exploit information on three cohorts of students. In Appendix Table A.7, I show the effect of the main specification presented in Table 5 for the three different cohort of students separately. Reassuringly also for the potential reverse causality concerns expressed in Section 4.4, results are not statistically different in the three cohorts, even if, since the number of observation decreases splitting the sample, estimates are noisier.[48]

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, in our sample of 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.14, 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.14, 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 Appendix D.

Finally, in the Appendix Table A.15, I consider the impact of self-reported gender bias. The impact on self-reported bias is generally small and in most specifications indistinguishable from zero. However, interestingly, controlling for reported bias, if anything increases the impact of implicit bias on gender gap in math performance. It supports the distinctiveness of implicit and explicit cognition (Greenwald et al., 1998) in the context of gender stereotypes of teacher.

[47] In Appendix E, I show the summary statistics for Italian teachers and I delve deeper into the role of Italian teachers.

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

6 Discussion of Potential Mechanisms

In this section, I discuss the mechanisms behind the negative impact of teacher bias on student achievement. I focus mainly on two aspects: self-stereotypes and interaction theory.[49] I use student survey data to analyze more deeply the former aspect, while for the latter I rely on the social psychology evidence on the interaction between teachers and pupils by gender. In Appendix G, I present a conceptual framework including both these aspects.

Self-Stereotypes

Self-confidence plays a crucial role in affecting performance, especially in gender-incongruent areas, such as female performance in math (Coffman, 2014). 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 beliefs that women have lower ability than men in math or are less suitable for a STEM career, females will 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).[50]

I find that biased math teachers activate negative self-stereotypes and induce females to believe that they are *worse at math* than what would be expected given their achievements. 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 teacher bias on math performance of female students.

Table 9 assesses the extent to which bias of teachers affect one's own assessment of ability, for a sample of around 800 students for whom I collected self-confidence measures, as described in section 3.2. I present results for self-stereotypes in math in Panel A, in reading in Panel B and

[49]There is a third theory 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, in appendix F we provide evidence that teachers favor female students in math grading, comparing blinded and no-blinded scores, as emerges in several other countries (Lavy and Sand, 2015; Terrier, 2015).

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

on average of all other subjects in Panel C. As shown in column 1, females are 9.4 percentage points less likely to consider themselves good at math (which corresponds to 11% percent lower probability than males), 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 is increasing. In particular, in classes assigned to teachers with one standard deviation higher bias, the gender gap in self-assessment increases by 4.5 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).

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 9, 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 achievements at the end of middle school. I find that gap in own assessment is reduced only by less than one third: teacher stereotypes have an additional impact on own assessment of math capabilities, on top of measured ability, that may have detrimental effects for investment choices in education and occupation.

In Appendix Table A.16, I show the result of the specification described with school fixed effects instead of class fixed effects (as in equation 2). Consistent with the results in Table 6, there is a negative impact of teacher bias on self-stereotypes of female students and no impact on male students. All results are robust to the inclusion of controls at pupil and teacher level and their interaction with student gender.

In Panel B and C of Tables 9 , I focus on the impact of math teacher bias on self-assessment respectively in Italian and all other subjects. Female students seem to compensate for the low confidence in math with higher self-assessment in Italian, the other main subject taught during middle school. There is no impact on other subjects. The effects are robust to the inclusion of controls at individual level (column 3 and 4) and at teacher level (column 4) and are coherent in both specification, including class and the school fixed effects (see Appendix Table A.16). Finally, in the last three columns of Panel B, I control for the standardized test score in Italian in grade 8: as expected, it does not affect the estimate since math teacher stereotypes do not impact gender gap in reading performance. A deeper analysis of the impact on reading test scores or of the gender gap of the Italian teacher is presented in Appendix E.

Interaction Theory

A second potential mechanism is related to the *interaction theory* (McConnell and Leibold, 2001): math teachers with higher gender bias may spend less time (in terms of either quantity or quality) interacting with girls, especially those performing poorly. Biased 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. However, we do not find evidence of higher achievement of these groups of students when exposed to a gender-biased environment. Unfortunately, I do not have measures of the "quality of interaction" between teachers and student by gender to directly test this mechanism.

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).[51].

Hence, biased teachers are more likely to fail to recognize talent of some students in math related fields and set a lower bar for their learning. Teachers' erroneous expectations may lead to a self-fulfilling prophecy: they may fail to recognize students' talent and therefore not encourage them to fulfill their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983). 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 males.

All these aspects suggest that gender-biased interaction between pupils and teachers is an important mechanisms behind the main results of this paper on the impact of teacher stereotypes on student achievements. They are also very important to understand the mechanisms through which self-stereotypes of students are activated when exposed to gender-biased teachers.

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 considering that on average in OECD countries less than 5 percent of 15-years-old girls are planning to pursue

[51]Using Italian data from INVALSI, I show in Appendix H 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.

a career in these fields compared to around 20 percent of boys according to 2015 PISA data. Social conditioning has a strong impact on development of skills and educational choices. This paper shows that the gender gap in math performance can be partially explained by teacher implicit bias. Females, especially those from disadvantaged backgrounds, are lagging behind when assigned to teachers with higher implicit stereotypes (as measured by an Implicit Association Test). Males, the group not ability-stigmatized in terms of math performance, are not affected by teacher bias. Teacher stereotypes affect high-school track choice, leading female students assigned to a teacher with higher implicit bias to be more likely to attend a vocational school. Furthermore, they foster low expectations about own ability and lead to underperformance in male-typed domains. Indeed, females 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 under-assess own ability and underperform fulfilling negative expectations about their achievements. Unconscious biases and 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 gap in math performance generated during middle school would be 35% smaller if no teachers had negative gender stereotypes (from 0.078 to 0.051 standard deviations). 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. Hence, one set of potential policies may be aimed at informing people about own bias or training them in order to assure equal behavior toward individual of ability-stigmatized groups and others. An alternative way to fight against the negative consequences of stereotypes is increasing self-confidence of female 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). More research is needed to further investigate the impact of both type of policies.

References

Alan, S., S. Ertac, and I. Mumcu (2017). Gender stereotypes in the classroom and effects of achievement. *Working Paper*.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2017). Revealing stereotypes: Teacher bias and immigrants performance. *Mimeo Bocconi Univeristy*.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. *Handbook of Labor Economics 3*, 3143–3259.

Banaji, M. R., B. A. Nosek, and A. G. Greenwald (2004). No place for nostalgia in science: A response to arkes and tetlock. *Psychological Inquiry 15*(4), 279–310.

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? *The 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.

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*.

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

Bordalo, P., K. B. Coffman, N. Gennaioli, and A. Shleifer (2016). Beliefs about gender. *NBER 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. Technical report, National Bureau of Economic Research.

Carlana, M., E. La Ferrara, and P. Pinotti (2017). Goals and gaps: Educational careers of immigrant children. *Mimeo Bocconi Univeristy*.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The 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? *The 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.

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. Technical report.

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. *The 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. *The 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? *The 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. *The 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.

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? Technical report, 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. *The Journal of Economic Perspectives 24*(2), 129–144.

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?

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. *The 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. *The 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*.

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

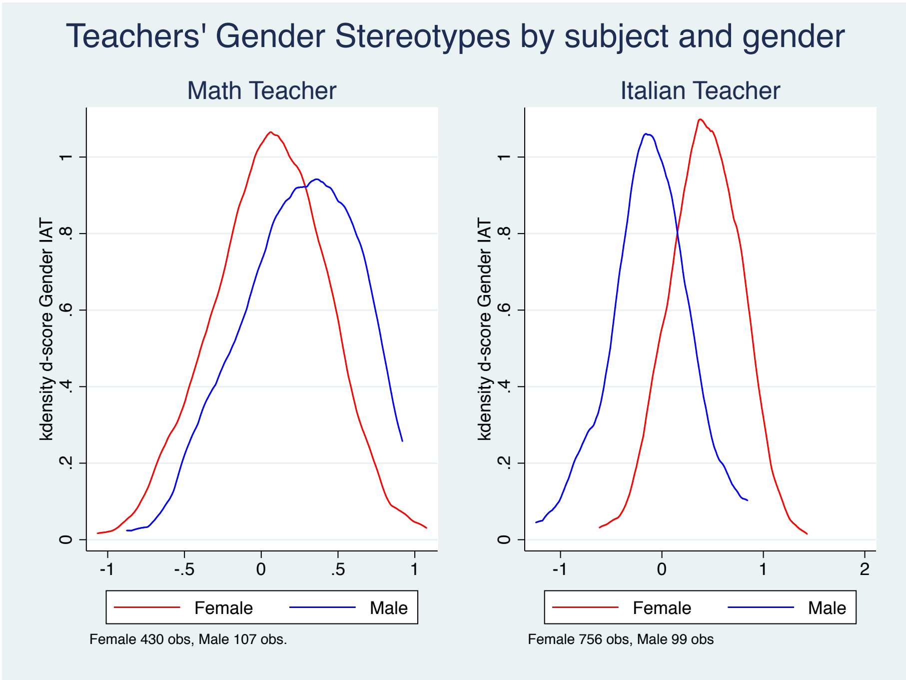

Figure 1: Teachers' Implicit Gender Bias (IAT measure) by gender and subject they teach

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. The graph provides evidence that teachers in gender-incompatible fields have stereotypes closer to zero.

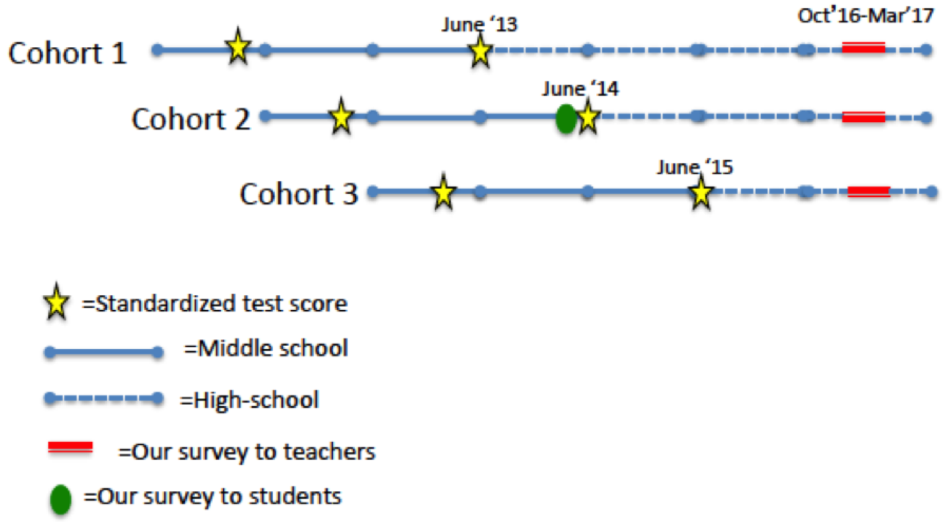

Figure 2: Timeline of main data available for students and teachers

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.

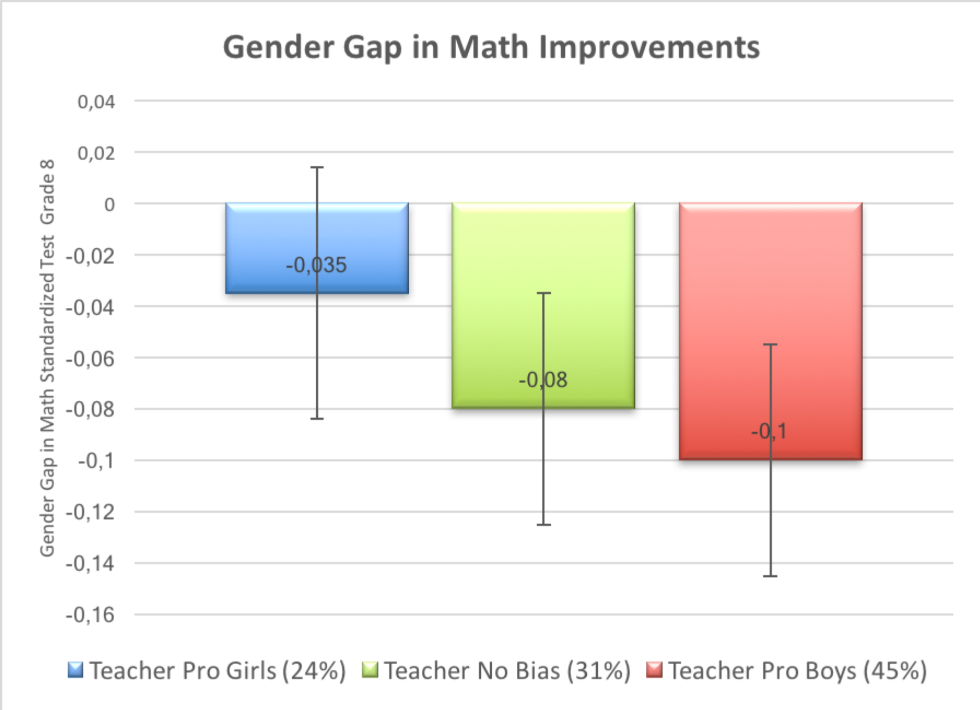

Figure 3: Effect of teacher bias on student math performance

Notes: This graph shows the effect of teacher stereotypes on student achievement. 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. The attitude of the teacher in associating fields with gender is considered "pro girls" if the score is lower than -0.15 (24% of teachers) and "pro boys" 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 absorbed.

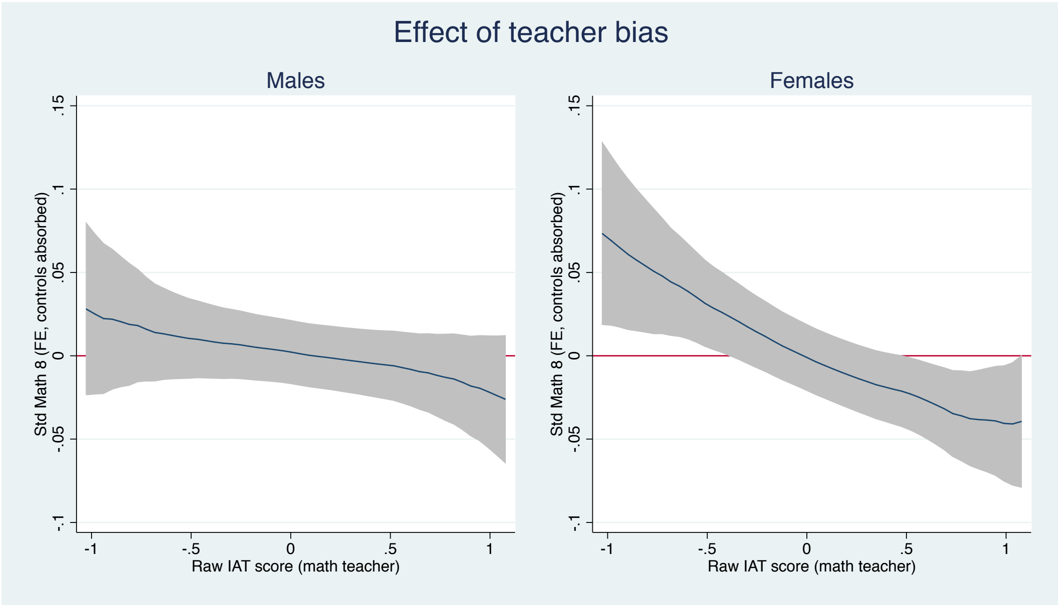

Figure 4: Effect of teacher bias on student math performance by gender

Notes: This graph shows the effect of teacher stereotypes on student achievement by gender. The variable 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.

Table 1: Summary Statistics from Math Teachers' Questionnaire

	Count	Mean	SD	Min	Max
Family and education					
Female	301	0.84	0.37	0.00	1.00
Born in the North	291	0.65	0.48	0.00	1.00
Age	290	51.90	8.38	31.00	66.00
Children	301	0.74	0.44	0.00	1.00
Number of children	215	1.84	0.80	0.00	5.00
Number of daughters	215	0.85	0.76	0.00	3.00
Low edu Mother	278	0.58	0.49	0.00	1.00
Middle edu Mother	278	0.29	0.46	0.00	1.00
High edu Mother	278	0.13	0.34	0.00	1.00
Advanced STEM	292	0.24	0.43	0.00	1.00
Degree Laude	256	0.17	0.37	0.00	1.00
Job characteristics					
Full time contract	285	0.92	0.28	0.00	1.00
Years of experience	287	22.94	10.79	3.00	48.00
Math Olympiad	292	0.19	0.39	0.00	1.00
Update Courses	292	0.94	0.24	0.00	1.00
Satisfy with teacher job	287	3.69	0.84	2.00	5.00
Implicit bias					
IAT Gender	301	0.09	0.37	-1.03	1.08
Self-reported explicit bias					
WVS Gender Equality	290	0.17	0.37	0.00	1.00
Gender Dif Innate Ability	280	1.51	0.76	1.00	3.00
Reason GenderGap: Interest for STEM	256	2.58	0.98	1.00	4.00
Reason GenderGap: Predisposition for STEM	241	2.12	1.03	1.00	5.00
Reason GenderGap: Low self-esteem	278	2.64	1.05	1.00	5.00
Reason GenderGap: Family support	278	3.14	1.08	1.00	5.00
Reason GenderGap: Cultural Stereotypes	279	2.15	1.16	1.00	5.00
Boys better in Invalsi	233	0.20	0.40	0.00	1.00
Girls better in Invalsi	233	0.32	0.47	0.00	1.00
Gender Equal in Invalsi	233	0.48	0.50	0.00	1.00
Observations	301				

Notes: First-hand data from teachers' questionnaire. We 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.2. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

Table 2: Summary Statistics of students by gender

	Males	Females	Diff.	se
Baseline characteristics				
Std Math grade 6	0.233	0.038	0.195***	(0.020)
Std Ita grade 6	0.085	0.218	-0.133***	(0.019)
Born in the North	0.849	0.854	-0.005	(0.007)
Born in the Center/South	0.027	0.030	-0.003	(0.003)
Immigrant	0.189	0.173	0.016	(0.008)
Second Gen. Immigrant	0.080	0.074	0.006	(0.006)
HighEduMother	0.456	0.453	0.003	(0.010)
Missing Edu Mother	0.212	0.211	0.002	(0.008)
High Occupation Father	0.169	0.174	-0.005	(0.008)
Medium Occupation Father	0.321	0.303	0.017	(0.010)
Missing Occupation Father	0.206	0.214	-0.008	(0.008)
Outcomes				
Std Math grade 8	0.194	-0.021	0.214***	(0.020)
Std Ita grade 8	-0.006	0.176	-0.182***	(0.020)
High-school Track: Scientific	0.304	0.208	0.096***	(0.010)
High-school Track: Classic	0.043	0.079	-0.036***	(0.005)
High-school Track: Other Academic	0.097	0.336	-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	0.132	0.148	-0.015*	(0.008)
Track recommendation: Scientific	0.164	0.110	0.054***	(0.008)
Track recommendation: Vocational	0.362	0.298	0.064***	(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	4698	4611		

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 3: Correlation between teachers' characteristics and Gender IAT Score

Panel A: Independent variables (background teachers' characteristics)

	Female (1)	Age (2)	HighMotherEdu (3)	Children (4)	Daughters (5)
Dep. Var.: Raw IAT score	-0.188**	0.016	-0.053	0.069	0.047
	(0.083)	(0.060)	(0.060)	(0.145)	(0.075)
Obs.	301	301	301	301	301
R^2	0.347	0.327	0.337	0.330	0.331

Panel B: Independent variables (cultural traits and beliefs)

	BornNorth (1)	WomenLFP (2)	WVSCityBorn (3)	WVSIndiv (4)	InnateAbility (5)
Dep. Var.: Raw IAT	-0.154**	-0.499**	0.399*	0.007	0.016
	(0.064)	(0.247)	(0.211)	(0.086)	(0.041)
Obs.	301	286	261	301	301
R^2	0.348	0.361	0.399	0.325	0.328

Panel C: Independent variables (education and teacher experience)

	Ad.STEM (1)	Laude (2)	FullContract (3)	Olympiad (4)	JobSatisfy (5)
Dep. Var.: Raw IAT score	-0.092	-0.034	-0.049	0.059	0.054*
	(0.076)	(0.075)	(0.153)	(0.087)	(0.032)
Obs.	301	301	301	301	301
R^2	0.332	0.326	0.327	0.311	0.336
School FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports OLS estimates of the correlation between math teachers' bias 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. School fixed effects are included in all regressions. The significance and magnitude of coefficients are not significantly impacted 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 "Ad.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, "Updates" captures whether teachers followed update courses in teaching during the academic year, "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,"ExplicitBias" is an index that summarizes explicit gender bias of teachers. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table 4: Exogeneity of assignment of students to math teachers with different stereotypes

Dependent Variable: Math Teacher implicit gender bias (standardized)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fem	0.007	-0.011	0.004	0.015	0.008	-0.018	0.220
	(0.013)	(0.022)	(0.025)	(0.016)	(0.013)	(0.132)	(0.239)
Fem*HighEduMother		0.036				0.044	-0.002
		(0.034)				(0.031)	(0.045)
HighEduMother		0.018				0.005	-0.009
		(0.027)				(0.025)	(0.029)
Medium Occupation Father			0.013			0.007	0.038
			(0.024)			(0.022)	(0.035)
Fem*Medium Occupation Father			0.020			0.008	0.076
			(0.036)			(0.033)	(0.060)
High Occupation Father			0.015			0.018	0.005
			(0.032)			(0.027)	(0.041)
Fem*High Occupation Father			0.006			-0.012	-0.032
			(0.041)			(0.038)	(0.059)
Fem*Immigrant				-0.035		0.005	0.097
				(0.038)		(0.040)	(0.076)
Immigrant				0.059^{**}		0.049^{*}	0.045
				(0.029)		(0.029)	(0.056)
Fem* Std Ita grade 6					0.005	-0.005	-0.005
					(0.015)	(0.015)	(0.026)
Std Ita grade 6					-0.009	-0.009	-0.016
					(0.013)	(0.013)	(0.017)
Fem*Std Mat grade 5							-0.002
							(0.025)
Std Mat grade 5							-0.005
							(0.016)
School,year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	Yes	Yes
Obs.	9309	9309	9309	9309	9280	9280	1649
R^2	0.412	0.412	0.412	0.412	0.419	0.489	0.723

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 math teacher level in parentheses; the number of clusters is 301 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 Mat grade 5" and "Std Ita grade 6" are the standardized test score in grade 5 in math and grade 6 in Italian respectively. 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. For 29 students we do not observe the test score in Italian in grade 6. 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 5: 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	-0.222***	-0.078***	-0.080***	-0.036	-0.024
	(0.019)	(0.014)	(0.014)	(0.032)	(0.103)
Fem*Bias Teacher			-0.027**	-0.028**	-0.037***
			(0.013)	(0.013)	(0.014)
Fem*Teacher Fem					-0.056
					(0.037)
Fem*North Math Teacher					0.008
					(0.030)
Fem*Advanced STEM Teacher					-0.041
					(0.031)
Std Math grade 6		0.723***	0.723***	0.697***	0.699***
		(0.012)	(0.012)	(0.013)	(0.013)
Constant	0.198***	0.028***	0.028***	-0.112***	-0.112***
	(0.009)	(0.007)	(0.007)	(0.023)	(0.023)
Gender Gap	-0.222	-0.078	-0.078	-0.082	-0.082
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9309	9309	9309	9309	9309
R^2	0.209	0.618	0.618	0.625	0.625

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 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 6: 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	-0.234***	-0.092***	-0.093***	-0.034	-0.020
	(0.022)	(0.015)	(0.015)	(0.033)	(0.107)
Fem*Bias Teacher			-0.022*	-0.024*	-0.032**
			(0.013)	(0.013)	(0.013)
Bias Teacher			-0.011	-0.011	-0.006
			(0.015)	(0.014)	(0.013)
Fem*Math Teacher Fem					-0.052
					(0.040)
Math Teacher Fem					0.061
					(0.041)
Fem*North Math Teacher					0.013
					(0.031)
Math Teacher born North					0.027
					(0.035)
Fem*Advanced STEM Teacher					-0.031
					(0.034)
Advanced STEM					0.026
					(0.034)
Std Math grade 6		0.716***	0.715***	0.687***	0.688***
		(0.011)	(0.011)	(0.012)	(0.012)
Gender Gap	-0.214	-0.077	-0.077	-0.081	-0.082
School, year FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes
Teacher Controls	No	No	No	No	Yes
Obs.	9309	9309	9309	9309	9309
R^2	0.136	0.576	0.577	0.585	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 301. 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 7: Estimation of the effet 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	-0.024	-0.020	0.033	-0.023	-0.050	-0.033
	(0.103)	(0.103)	(0.112)	(0.103)	(0.104)	(0.104)
Fem*Bias Teacher	-0.037***	-0.049**	-0.070***	-0.036**	-0.040**	-0.065**
	(0.014)	(0.021)	(0.027)	(0.015)	(0.016)	(0.031)
Fem*Bias T*HighEduM		0.022				
		(0.028)				
Fem*Bias T*Top tercile Math6			0.100***			
			(0.035)			
Fem*Bias T*Middle tercile Math6			0.011			
			(0.035)			
Fem*Bias T*Immigrant				-0.011		
				(0.038)		
Fem*Bias T*Extended School Day					0.012	
					(0.026)	
Fem*Bias T*Same Math Teacher						0.031
						(0.035)
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.	9309	9309	9309	9309	9309	9309
R^2	0.626	0.626	0.627	0.626	0.626	0.626

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 301. The number of fixed effects (classes) is 548. 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 daughthers, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, 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 8: Estimation of the effet of teachers' gender stereotypes on track choice- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A- Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.094***	-0.048***	0.025	0.171*	0.014	-0.009	0.020	0.016
	(0.012)	(0.011)	(0.019)	(0.092)	(0.009)	(0.010)	(0.023)	(0.071)
Fem*Bias Teacher		0.009	0.008	0.001		0.023**	0.020**	0.020**
		(0.012)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Fem* Teacher Fem				-0.032				0.034
				(0.029)				(0.022)
Std Math grade 6		0.178***	0.159***	0.159***		-0.104***	-0.091***	-0.091***
		(0.008)	(0.008)	(0.008)		(0.007)	(0.007)	(0.007)
Constant	0.299***	0.242***	0.106***	0.108***	0.141***	0.174***	0.207***	0.205***
	(0.006)	(0.006)	(0.015)	(0.015)	(0.005)	(0.006)	(0.016)	(0.016)
Mean Y for Fem	0.205	0.205	0.205	0.205	0.155	0.155	0.155	0.155
Obs.	8463	8463	8463	8463	8463	8463	8463	8463
R^2	0.113	0.214	0.233	0.236	0.119	0.190	0.208	0.211
	Panel B- Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.045***	-0.019**	0.033**	0.016	-0.059***	-0.110***	-0.125***	-0.059
	(0.010)	(0.009)	(0.015)	(0.081)	(0.013)	(0.011)	(0.024)	(0.092)
Fem*Bias Teacher		0.001	-0.000	-0.007		0.018*	0.018*	0.024**
		(0.009)	(0.009)	(0.009)		(0.010)	(0.010)	(0.011)
Fem* Teacher Fem				-0.053**				0.024
				(0.025)				(0.036)
Std Math grade 6		0.126***	0.113***	0.113***		-0.246***	-0.217***	-0.217***
		(0.009)	(0.009)	(0.009)		(0.008)	(0.008)	(0.008)
Constant	0.156***	0.129***	0.059***	0.059***	0.376***	0.428***	0.518***	0.517***
	(0.005)	(0.004)	(0.011)	(0.011)	(0.006)	(0.006)	(0.017)	(0.017)
Mean Y for Fem	0.110	0.110	0.110	0.110	0.317	0.317	0.317	0.317
Obs.	7086	7086	7086	7086	7086	7086	7086	7086
R^2	0.152	0.238	0.249	0.251	0.150	0.362	0.389	0.391
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 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 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, children and daughters, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, 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 9: Estimation of the effet of teachers' gender stereotypes on self-stereotypes- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)							
Fem	-0.094***	-0.067**	-0.093	0.174	-0.053*	-0.074	0.195
	(0.029)	(0.028)	(0.065)	(0.188)	(0.028)	(0.065)	(0.200)
Fem*Bias Teacher		-0.045**	-0.049**	-0.066**	-0.030	-0.033	-0.052*
		(0.021)	(0.022)	(0.030)	(0.021)	(0.023)	(0.030)
Std Test Math		0.138***	0.135***	0.136***	0.157***	0.151***	0.148***
		(0.024)	(0.023)	(0.024)	(0.023)	(0.023)	(0.024)
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**	0.057**	0.045	0.166	0.047**	0.035	0.135
	(0.023)	(0.023)	(0.048)	(0.215)	(0.021)	(0.046)	(0.203)
Fem*Bias Teacher		0.038**	0.038**	0.026	0.038**	0.039**	0.029
		(0.018)	(0.019)	(0.022)	(0.017)	(0.019)	(0.021)
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	0.019	0.021	-0.213	0.016	0.018	-0.219
	(0.027)	(0.029)	(0.062)	(0.224)	(0.028)	(0.062)	(0.227)
Fem*Bias Teacher		-0.014	-0.015	-0.020	-0.018	-0.020	-0.025
		(0.023)	(0.024)	(0.027)	(0.024)	(0.024)	(0.027)
Constant	1.672***	1.689***	1.674***	1.681***	1.687***	1.670***	1.676***
	(0.014)	(0.016)	(0.041)	(0.041)	(0.015)	(0.040)	(0.040)
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 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.