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

Michela Carlana†

June 2017

JOB MARKET PAPER

(This is a preliminary draft. Please do not cite or distribute without permission of the author.)

LATEST VERSION HERE AND APPENDIX HERE

Abstract

We explore the influence of teachers' gender stereotypes in affecting student performance, combining administrative data and original first-hand questionnaires on both teachers and students. We find that gender gap in math performance increases when students are quasi-randomly assigned to teachers with higher implicit bias (as measured by an Implicit Association Test). The effect is driven by students from disadvantaged background and by lower performance of female students, while there is no effect on males. Teacher bias has a substantial impact on own assessment of math ability. Our results show that biased teachers activate negative self-stereotypes on female students only in male-typed domains. The findings are consistent with a model of stereotype whereby ability-stigmatized groups underperform fulfilling negative expectations about their achievements.

JEL: J16, J24, I24.
Key words: gender gap, math, teachers, stereotypes, self-stereotypes.

*I am grateful to Alberto Alesina, Eliana La Ferrara and Nicola Gennaioli for insightful comments and encouragement. I thank for their useful suggestions Ingvild Almas, Thomas Le Barbanchon, Pamela Giustinelli, Joseph-Simon Goerlach, Selim Gulesci, Andreas Madestam, Paolo Pinotti, Jonathan de Quidt, Dan-Olof Rooth, David Stromberg, Jackob Svensson, Guido Tabellini, Anna Tompsett and all seminar participants of IIES Brownbag, Bocconi F4T Brownbag. Giulia Tomaselli, Elena De Gioannis, Sara Spaziani and Tommaso Coen 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 grateful to the Italian Ministry of Education (Statistics Office), INVALSI and all principals and teachers of schools involved in the research for giving me access to data used in this paper.

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

1 Introduction

Over the last century, the narrowing of gender differences has been impressive, up to a reversal of the gap in school attainment in many OECD countries. Despite that, gender stereotypical beliefs are pervasive and deeply-held in most societies (Alesina et al., 2013). Women are believed to be worst than men in highly profitable fields as mathematics, engineering and technology, even controlling for measured ability (Reuben et al., 2014). Gender stereotypes are overgeneralized and amplified representations of real differences, but they often hold a *kernel-of-truth* (Bordalo et al., 2017). Gap in math performance between males and females in both developed and developing countries is still far from being closed[1]. To the extent that gender stereotypes are internalized directly in the development of self-concept or influence investment choices, these cultural beliefs may have causal influence on life-outcomes of individuals, shaping educational and occupational careers. In this paper, we explore whether exposure to stereotypes can causally affect math achievements of boys and girls.

We focus in particular on the influence of teachers' gender stereotypes in affecting student performance, combining administrative data and original first-hand questionnaire on both teachers and students in Italy. We find that gender gap in math performance increases when students are quasi-randomly assigned to teachers with higher bias (as measured by an Implicit Association Test). The effect is driven by students from disadvantaged backgrounds and by lower performance of female students, while male students are not affected by implicit stereotypes. Teacher bias has a substantial impact on own assessment of math ability, as measured by detailed information collected through an original student questionnaire. Our results show that biased teachers activate negative self-stereotypes on female students only in male-typed domains. Furthermore, we also provide evidence that teachers' implicit bias is correlated with their high-school recommendation to students and it has an influence on high-school track choice.

The findings are consistent with a model of stereotype whereby ability-stigmatized groups underperform fulfilling negative expectations about their achievements. Teacher bias fosters low expectations about own math ability and, through this mechanism, underperformance of individuals vulnerable to the gender stereotype[2]. We collect data to provide evidence about this channel, which is most likely not the only one in place. Teacher bias may simultaneously increase anxiety of female students, but it may also influence directly the time and quality of

[1] According to a meta-analysis performed on 100 studies in several countries, gender gaps in mathematics are around 0.29 standard deviation in high-school (Hyde et al. (1990)).

[2] 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). This conceptual framework is related for example to the model of Coate and Loury (1993); Lundberg and Startz (1983), in which bias impede skills development. Furthermore, even if we do not explicitly refer to the stereotype threat model, the implications are similar to the economic model behind it described in Dee (2014).

interaction between teachers and students[3].

Economists have mainly focused on two forms of *conscious* discrimination: "taste-based" discrimination (Becker et al., 1957), resulting from a sort of animus toward members of the out-group, and "statistical" discrimination (Phelps, 1972; Arrow et al., 1973), coming from rational expectations and imperfect information. Human behavior was regarded as motivated by rational thought, but many exceptions are recognized and included in models of stereotypes (Tversky and Kahneman, 1975; Bordalo et al., 2017). Recently, the economic literature has underlined the benefits from interacting with social psychologists and considering *unconscious* bias in studying discrimination (Guryan and Charles, 2013; Bertrand and Duflo, 2016). Gender stereotypical beliefs can operate even without awareness, explicit endorsement or intention to harm the stigmatized-group (Nosek et al., 2002). It can operate through subtle forms of "encouragement" to consider own math achievements as the result of talent for boys and effort for girls[4]. Hence, the message sent may induce females to believe "math is a male field" and that they cannot make it when math will get more complicated[5]. Furthermore, a teacher may systematically call-out a male student to solve the most difficult math problems and female students to solve easier one. This differential treatment may enhance the association of math with males, ceteris paribus, and differentially affect the self-perception of own math ability by gender, even controlling for measured performance.

Building on the recent development of economic literature studying discrimination, we collect teachers' gender bias using the Implicit Association Test (IAT), a measurement tool developed around twenty years ago and widely used by social psychology (Greenwald et al., 1998). This computer-based test measures the relative strength of association between pairs of concepts: names typically associated with boys or girls and subjects related to scientific or humanistic fields. When completing the test, participants are asked to categorize words as fast as possible: for instance, all female names and scientific subjects on the left of the screen and male

[3]The connection between *interaction* and bias has been described for instance by McConnell and Leibold (2001). Furthermore, biased teachers may dislike females and assigned to them unpleasant tasks, as hypothesized by *animus* theory, even if this mechanism seems the least likely in our context. However, we do not have data to explore the influence of these additional potential mechanisms.

[4]We have information on a sample of almost 20000 students in grade 6 collected by the Italian National Evaluation Center (INVALSI) on the reasons why students are performing well in both math and reading. In particular, students are asked: *"You solved all exercises in a math exam correctly. Why has it happened?"* and *"You need to tell your classmates about a text you read. You do it in a clear and precise way and everybody follows what you are saying. Why has it happened?"* Students can choose among different potential answers: "I was helped", "I was lucky", "It was easy", "I am good" and "I have exerted a lot of effort". Female students are 50% less likely to say "I am good" in math compared to reading, while males are only 13% less likely to believe they are good in math compared to reading. Furthermore, females are also 49% less likely than males to say they are good in math as answer and the effect is even stronger when conditioning to math standardized test scores. For more details, see Table A.1.

[5]In a potentially extremely simplified example, a teacher may say "You are talented in math. If you study a bit harder, you will get a top grade!" to a male student and "Good, I can see you have studied hard!" to a female student who got the same grade, exerting the same level of effort.

names and humanistic subjects on the right. The IAT score is determined by the difference in reaction time (*Male+Scientific, Female+Humanistic* vs. *Female+Scientific, Male+Humanistic*). The underlying assumption is that the responses are faster and more accurate when names and fields are closely associated by our brain as compared to when they are not (or less closely associated). IAT scores have been found to correlate with many outcomes in the real-world and in the lab[6] and they are often correlated but distinct from explicit and self-reported bias (Lane et al., 2007).

The schooling context is particularly interesting for studying the impact of gender stereotypes. Gender gap in math performance varies substantially across countries and throughout the educational career, raising the question on how strong the influence of the schooling environment is, even on top of parents and society impact. The difference in math performance between boys and girls is wider in those countries with low women empowerment and higher implicit gender bias (Guiso et al., 2008; Nosek et al., 2009) [7]. Furthermore, there are no (measurable) differences upon entry to school (Fryer Jr and Levitt, 2010), but the gender gap gets larger across the educational life course of students[8]. The ground lost by girls relative to boys in math during school years may be due to several aspects, as family background, investments to school environment, evolutionary foundation and biological differences in brain functioning[9]. However, the two facts related to the cross-country difference in gender gap and enhancement of discrepancies throughout the career of students points toward an important role of social-conditioning in influencing math performance of boys and girls.

Despite a broad literature has tried to investigate the causes of gender gap, the host of potential cultural explanations analyzed does little to explain this gap. For instance, Fryer Jr and Levitt (2010), using a nationally representative sample in the USA, find little empirical support for potential factors as parental expectations and time spent with children doing math-related activities, less investment by girls in math or biased tests. On the other side, Niederle and Vesterlund (2010) argues that gender differences in competitiveness may have distortion effects and exaggerate the advantage of males, especially in the right tail of the distribution of test scores. Bharadwaj et al. (2016) analyze a detailed dataset from Chile and find that the gender

[6]For instance, Nosek et al. (2009) shows the correlation between gender IAT scores with gender gap cross-countries and Reuben et al. (2014) shows in a lab experiment that higher stereotypes (measured by gender IAT) predict employers' bias expectations against female math performance. Also race IAT scores (using "Black and White" names and "Good and Bad" adjectives) have been shown to correlate with real-world outcomes (Greenwald et al., 2009; Rooth, 2010; Glover et al., 2017).

[7]Dickerson et al. (2015) shows that in Africa the gender gap in math performance is strongly correlated with fertility rates, even controlling for GDP.

[8]Coherently with the international evidence, in the Italian context for instance, Contini et al. (2017) shows that the gender gap in achievement test scores is 0.09 standard deviations in grade 2 and 0.28 standard deviations in grade 10.

[9]For instance, Baron-Cohen (2003) elaborated the "empathizing-systemizing theory" according to which there are evolutionary differences among the two genders: females are stronger empathizers and males are stronger systemizers.

gap is sizable even within twins, suggesting that the effect is not driven by family characteristics. Furthermore, controlling for the gender of the math teacher does little to close the gap. Teacher attitudes seem to matter: Alan et al. (2017) investigates the impact of self-reported gender stereotypes in Turkey finding a significant relation between girls' performance in both math and reading. Lavy and Sand (2015) and Terrier (2015), respectively in Israel and France, analyze the impact of bias in teachers' assessment, as measure by gender differences in blind and no-blind test-scores within the class, and find a strong correlation with future performance of students. Compared to other measures of teachers' bias, the Implicit Association Test has two main advantages. First, with respect to self-reported information, it does not suffer by social desirability bias. Second, with respect to bias in grading, teachers observe additional gender specific factors when they grade students, which may be correlated with future performance of students [10].

Girls outperform boys in reading and in other dimensions of school attainments. So why should we care about gender gap in math? Several studies have shown that math test scores are good predictors of future occupation and earning of individuals (Altonji and Blank, 1999). Gaining a better understanding of the reasons behind the emergence of gender gap in math skills is of first-order importance to explain the enduring gender gap in performance and the underrepresentation of women in leadership position and among science, technology, engineering, and math (STEM) workforce.

The paper proceeds as follows. Section 2 explains the conceptual framework and proposes a stylized model of the influence of teachers' stereotypes on students' math performance. Section 3 explains the setting analyzed - the Italian schooling System. Section 4 describes the data available on both students and teachers and provides evidence in favor of quasi-random assignment of students to teachers with different implicit bias. Section 5 presents the main results of the paper, showing that gender gap increases during middle school in those classes assigned to more biased teacher, as well as the gap in self assessment of own math ability and the probability of attending a more demanding high-school. Finally, Section 6 concludes. All supplementary material is provided in the Appendix.

2 Conceptual Framework

Gender stereotypes are deeply rooted in most societies and therefore female students are potentially vulnerable of being judged by the predicament that "girls are not good at math"[11]. There

[10] In (Carlana, 2017), I provide evidence that in Italy there is low correlation between classes taught by the same teacher in the measure of bias in grading.

[11] I do not contribute to the debate on how stereotypes are formed. They may come from rational expectations as in a model of statistical discrimination (Arrow et al., 1973; Phelps, 1972) or from the representativeness of types as in Bordalo et al. (2017).

are different mechanisms through which teacher stereotypes can affect student performance. First, if the student perceives higher bias toward own group, she may decrease assessment of own ability in math. The major role for self-stereotyping in shaping beliefs about own ability has been recently uncovered by Coffman (2014) and Bordalo et al. (2016). Girls may believe that both own signal of ability and the signal received by teachers carry relevant information. However, if the signal received by teachers is biased given the documented and widespread belief that women have lower ability than men in math, females will develop a lower self-assessment of own ability in the scientific field and potentially invest less in their STEM education. Self-beliefs are one of the potential channels through which stereotypes of the society have an impact on gender gaps in math performance. This theory is consistent with the idea of stereotype threat (Steele and Aronson, 1995) developed by social psychological literature, 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). Despite the rich literature in social psychology about stereotype threat since 1990s, only recently economists have analyzed directly this phenomenon. One of the first step taken in this direction has been Fryer et al. (2008) that 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.

A second potential mechanism is related to the *interaction theory* (McConnell and Leibold, 2001): math teachers with higher gender bias may feel uncomfortable in interacting with girls, pay less attention in finding the adequate method to help them learning or simply be worried to seem prejudiced. Alternatively, they may believe that it is less productive to spend time explaining math to girls since they have lower returns compared to males or will not attend a STEM track in any case (especially those not in the top tercile of the distribution). Girls may realize that and spend less time in studying math. For instance, Glover et al. (2017) find that managers of French grocery stores with higher implicit race bias interact less with minorities putting them less pressure on exerting high effort at work. They are also less likely to assign minority employees to unlikable cleaning duties, most likely to avoid interaction with them or not to seem prejudiced.

Finally, teachers with higher bias may simply dislike female students (*animus theory*), treat them badly and give 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 and in general teacher (no-blind) grading vs. blind grading is in favor of girls compared to boys (as in most countries).

We develop a simple conceptual framework to analyze the impact of teachers' gender stereotypes on effort of students [12], as mediated by student perception of own ability. In this simple

[12] In the empirical application,we measure improvement in achievement test scores and not effort. The reason is

framework, students choose effort and individual's utility can be represented as $u_i = \theta_i k(a_i, e_i) - c(e_i)$, where k is the benefits function, which depends on ability (a_i) and effort (e_i), and the cost c is paid according with the level of effort e exerted by individual i. The component θ_i introduces an exogenous heterogeneity and it captures observable difference across individuals in the returns to effort and ability. Dee (2014) presents an economic model of stereotype threat that is strongly related to the one presented in this paper. In the empirical counterpart of this model, we observe performance (P) and not directly effort (e), but we assume for simplicity that the derivative of performance with respect to ability is positive ($P_e > 0$) and in this section we focus on the choice of effort of students.

We extend this framework to capture belief formation about own ability level and the influence of teachers' gender bias through self-beliefs. We define x_i as the vulnerability to the stereotype that "girls are not good at math", which is based on observable characteristics of individual i. For instance, a female compare to male student is assumed to be more vulnerable to this predicament since it is part of the ability-stigmatized group. Student perception of teacher's stereotypes (ρ_{it}) depends on the vulnerability of the student i and the stereotype of the teacher t. This simple framework is flexible enough to capture heterogeneous perception of teachers' bias among classmates. For instance, boys may not perceive gender stereotype in math since they do not belong to the stigmatized group. The self-assessment of ability (α) is influenced by student's perception of teachers' stereotypes[13]. We assume that students' beliefs about own ability is a decreasing function of perceived teachers' stereotype, i.e. $\alpha_\rho \leq 0$[14].

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

$$u_{it} = \theta_i k(\alpha_{it}, e_{it}) - c(e_{it}) \tag{1}$$

where $\alpha_{it} = f(\rho_{it})$, u is differentiable and the sufficient conditions for a local interior maximum hold. In this model, we do not introduce parametric assumptions on the utility function.

We are interested in how the optimal level of effort of students varies according with the perceived stereotype of the teacher. The model implies that:

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

The second order condition for a relative maximum implies that the second order derivative

twofold: first (and crucially) we do not observe it directly, second it is not clear which is the best measure of effort: hours of study, are not necessarily a good measure since the quality of time use is also essential in the learning process.

[13]We do not model the formation of stereotypes that may come either from rational beliefs as in a statistical discrimination model (Arrow et al., 1973; Phelps, 1972) or from the representativeness of types as in Bordalo et al. (2017).

[14]Adding structure to this simple model, we may model α as the posterior update of the own ability after receiving a biased signal from teachers.

must be negative and therefore the denominator in equation (2) must be positive. Furthermore, we assumed that $\alpha_\rho \leq 0$. Hence, the optimal level of effort with respect to the perceived stereotype (e^*_ρ) depends on the complementarity or substitutability of effort and perceived ability ($k_{e\alpha}$). Suppose that effort and ability are complement (i.e. $k_{e\alpha} > 0$), then students who perceive higher teacher' gender stereotypes will decrease the level of effort in equilibrium ($e^*_\rho \leq 0$). The opposite is true if effort and ability are substitutes. As suggested also by Dee (2014), the latter case is likely for instance if individuals of the stigmatized group consider the stereotype strongly improper and react with an "I'll show you are wrong" attitude. In the context of gender stereotypes, it would imply that talented female students may increase the level of effort when they interact with teachers with stronger bias in order to disprove the negative belief [15].

Teacher stereotypes may influence the level/quality of effort of teachers in interacting respectively with females and males. Teachers investment may have a direct impact on perceived ability of students and their effort level and the introduction of this aspect goes beyond the scope of this simple framework. In the the empirical analysis, we do not observe gender specific investment or interaction in the classroom between professors and students, but we focus on the following thoughtful experiment. Assume two students have the same vulnerability to the stereotype ($x_i = x_j$), but they are quasi-randomly assigned to two different teachers, respectively with stereotypes s_{t_i} and s_{t_j}, such that $s_{t_i} < s_{t_j}$. Then, the perceived stereotype of student i is lower than the perceived stereotype of student j, i.e. $\rho_{it_i} \leq \rho_{jt_j}$. We can test the implication on own assessment of ability of the students and we expect $\alpha_{it_i} \geq \alpha_{jt_j}$. We will analyze how improvements in achievement test scores are affected by perceived stereotypes. If effort and ability are complement (as we may expect on average), then according with our conceptual framework, the optimal level of effort and therefore performance of the student decreases with perceived teachers' stereotypes.

3 Setting

We study the influence of teachers' gender stereotype exploiting an original dataset we created combining administrative information and original first-hand questionnaire, on both Italian students and teachers. Italy is a country with low labor market participation of women: the average employment rate of females in 2016 was 47 percent, with substantial geographic variations within the country [16]. We focus on a sample of middle schools located in the North of Italy

[15] Fryer et al. (2008) finds that stereotype threat do not affect women performance in math in a lab experiments where subjects recruited for the experiment were students of University of Chicago, one of the top US universities. In the light of our conceptual framework, the positive but statistically insignificant average effect of the stereotype threat is not surprising if abound half of these female students are characterized by an "I'll show you are wrong" attitude ($k_{e\alpha} > 0$).

[16] The employment rate of women was 31 percent in the South and 58 percent in the North, where the employment rate of men was respectively 55 and 74 percent (Source: Istat).

(Milan, Brescia, Padua, Genoa and Turin), an area close to the average OECD country in terms of women labor marker participation and empowerment[17]. Students at the beginning of middle school in grade 6 are assigned to classes and they stay with the same peers for three years, unless they are retained or they transfer to a different school[18]. The general class formation criteria are established by an Italian law and details are specified by each school council[19]. The general criteria mentioned are equal allocation of students across classes according with gender, nationality, disability, socio-economic status and ability level (as reported by the elementary school). In this research, we also collect new information directly from the principal on how classes are formed and they report the homogeneity across classes and heterogeneity within classes are the most important aspect[20] and this intention of the principal is confirmed by data analysis in our sample[21].

Teachers are assigned to schools by the Italian Ministry of Education and they are all paid equally for the same amount of hours. Teachers' allocation across school is determined by seniority: teachers tend to move away from disadvantaged backgrounds, where less experienced teachers are left (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 6 to 8 grade. Every week, students spend at least 6 hours with the math teacher and 5 hours with the Italian teacher[22].

Standardized test score in math and reading are administered in grade 2, 5, 6[23], 8 and 10 by the National Center for the Evaluation of the Italian Educational System (Istituto Nazionale per la VALutazione del SiSTEMa educativo di Istruzione e formazione, INVALSI). The achievement test score of grade 8 is the highest stake among these test scores, since it will affect 1/6

[17]The average employment rate in 2016 in OECD countries is 59.5 percent (Source: OECD).

[18]The Italian schooling System is offered free to all children, regardless of nationality, from age 3 to age 19. The schooling System is organized in a pre-primary school lasting three years, primary school lasting five years, a middle school lasting three years and an high-school of five years. Higher education is offered by universities. Education is compulsory from age 6 to age 16.

[19]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. The information on class formation criteria is available on a formal document on the website of each school of our sample and it is called the "Piano dell'Offerta Formativa".

[20]Among 81 principals, 64 percent consider the heterogeneity across classes in the ability level as "Extremely important" and 33 percent as "Important". The heterogeneity across classes in the socio-economic status is considered as "Extremely important" by 60 percent of principals and "Important" by 29 percent. The equal allocation of immigrant across classes is considered "Extremely Important" by 25 percent of principals and as "Important" by 38 percent.

[21]Using a Pearson Chi-squared test based on baseline characteristics, I find that only 7.2% have a p-value equal or lower that 5 percent. Furthermore, an analysis of Ferrer-Esteban (2011) shows that there is heterogeneity across classes within school in family background almost exclusively in the South of Italy. Our sample includes only cities in the North with the highest level of homogeneity in class formation.

[22]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

[23]This test score was administered only up to 2012-13.

of the final score of students at the end of middle school. However, this grade has no relevant impact for the enrollment in high-school or for the future education career of students. The tests are presented to all students as ability tests, thus making the gender stereotype in math potentially relevant. Students do not write their name on the exam, which is not corrected by their own teachers of the same subject, and will not know officially their result on the exam for all achievement tests except in grade 8. Finally. students receive grades by teachers at the end of each semester, which may be affected not only by performance, but also potentially other factors as diligence, effort and improvements over time. Grades are given in a scale up to 10, where the pass grade is 6.

The Italian education System is characterized by stratification of students after grade 8 into three different tracks: academic oriented high-school ("liceo"), technical high-school and vocational high-school. Students are free to choose the high-school they like the most, with no restriction based on grades or ability, and Giustinelli (2016) has shown that child's enjoyment of the curriculum is one of the most valued attribute. Teachers give a non-binding track recommendation to families with an official letter sent to children's home.

4 Data and Descriptive Statistics

In September 2016, we invited 156 middle schools to take part in a broad research project regarding "The role of teachers in high-school track choice"[24], out of which 103 accepted to take part into our study [25].

We use four sources of data: administrative information from the Italian Ministry of Education and from the National Center for the Evaluation of the Italian Educational System (INVALSI), teacher survey data and student survey data. The former provides information on gender, place of birth, high-school track choice, grades given by teacher and their track recommendation to students. INVALSI provides information on standardized test scores and family background. Finally, we collected detailed information on teachers, including the Gender Implicit Association Test (IAT), and on students' self-assessment of own ability.

[24]The research project was considered consistent with the ethical principles applied by Bocconi University and it was approved by the Ethics Committee of Bocconi University on 14th September 2016.

[25]The sample was initially designed including all schools of Milan, Brescia, Padua, Genoa and Turin with more than 20 immigrants in the scholastic year 2011-12 enrolled in grade 6. Among these 103 schools, the enumerators were able to collect the authorization of the principal to match data from students and teachers of 91 schools. In some cases, despite the collaboration of teachers, the principal preferred not to give access to data (2 cases) or we were not able to contact them directly in time before submitting our data application to the Ministry of Education and to the National Center for the Evaluation of the Italian Educational System (INVALSI).

4.1 Teacher Survey

From October 2016 to March 2017, we conducted a survey of around 1.400 math and Italian teachers in those schools that agreed to take part into the research project. The questionnaire was administered directly by enumerators using tablets in a meeting held in school buildings, in most of the cases in the early afternoon. Participants agreed on being part of the survey and signed an informed consensus, in which it was explained that the survey was part of a research project aimed at analyzing the role of teachers in affecting students' track choice. There was no direct reference specifically to gender bias. The time to complete the whole survey was around 30 minutes and teachers did not receive compensation for it. Among all math and Italian teachers working in the schools involved in this research, around 80 percent completed our survey[26].

Detailed information about the IAT test are provided in Appendix B. After IATs, we collected detailed information about family background of teachers (age, parents' education, age and sex of children, place of birth, ...) 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, number of upgrade courses done in the previous academic year, ...). We also collected information about the weight in a scale from 1 to 5 of different 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)[27], which factors teachers consider as the most important in grading (performance in oral and written exams, homework and diligence, attentive attitude) and in track recommendation (performance at school, interests of students, education of parents, interest in school activities of parents and economic resources of the family). We also collected information about explicit gender bias, as for instance beliefs about gender differences in innate math ability and the Word Value Survey question: *"When jobs are scarce, men should have more right to a job than women"* (with potential options "*Agree*", "*Neither Agree nor Disagree*", "*Disagree*"). Finally, we asked teachers to assess the average performance of their students in the standardized test score, by gender, and to reveal how sure they were about their answer. We obtained information on the classes in which they have been teaching from 2011 and we double checked all these information using data provided directly by schools and their website.

We build a measure of "Reported Gender Bias" of the teacher using the information obtained from the assessment on the reasons for the gender gap in the choice of the STEM track, beliefs about innate ability and access to labor marker. Factor analysis is performed using polychoric correlation matrix since variables are ordinal and separately for math and Italian teachers. Factor

[26]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 sense of the survey.

[27]The question asked is the following: *"Female students are less likely to attend a scientific track during high-school, controlling for math grades. Why do you think this is the case?"*

loadings are presented in Table A.2. 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.

4.1.1 Implicit Gender Bias

Economists have mainly focused on two form of *conscious* discrimination: "taste-based" discrimination (Becker et al., 1957), resulting from a sort of animus toward members of the out-group, and "statistical" discrimination (Phelps, 1972; Arrow et al., 1973), coming from rational expectations and imperfect information. Human behavior was regarded as motivated by rational thought, but many exceptions are recognized and included in models of stereotypes (Tversky and Kahneman, 1975; Bordalo et al., 2017). Recently the economic literature has underlined the benefits from interacting with social psychologists and considering *unconscious* bias in studying discrimination (Guryan and Charles, 2013; Bertrand and Duflo, 2016). In particular, social psychology suggests that both explicit and implicit attitudes affect individuals' actions. Deeply-rooted cultural and societal norms may affect how individuals behave toward specific groups and their behavior can also contradict their explicit views or self-interest (Greenwald and Banaji, 1995; Bertrand et al., 2005). Hence, gender stereotypical beliefs can operate even without awareness or intention to harm the stigmatized-group. Examples abound. Parents may choose gender-specific toys that induce a differential development in children (as building and construction toys for boys and kitchen toys and dolls for girls) or steering their children toward differential educational choices[28]. Similarly, math teachers may unintentionally challenge more male students on average asking to reply out loud to more difficult questions in class or by being more generous in assessment and in the verbal interaction toward females, setting a lower bar for their achievements.

Fully acknowledging the importance of considering implicit bias in economic literature studying discrimination, we collect teachers' gender bias using Implicit Association Test (IAT), a measurement tool developed around twenty years ago and widely used by social psychology (Greenwald et al., 1998; Lane et al., 2007). The idea underlying the test is that the easier the mental task, the faster the response production and the fewer the errors made in the process[29]. 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

[28]Indeed, according to the answers in our teacher survey reported in Table 1, the most important reason for the lower enrollment of females compared to males with the same math ability in scientific high-school track is parents' recommendation toward alternative choices.

[29]This concept was initially developed by Donders (1868). Donders was very optimistic on 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.

(as for instance gender and scientific/humanistic subjects). We administered the test using touch screen tablets. Subjects were presented with two sets of stimuli: (1) typically Italian names of females (e.g. Anna) and males (e.g. Luca), (2) 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 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. To calculate the score, two types of blocks are used [30]: 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). The order of the two types of blocks is randomly selected at individual level. The IAT score shows up as differential in response time between order compatible and incompatible blocks. It provides an index of the relative strength of association between *Male+Scientific* (and *Female+Humanistic*) vs. *Female+Scientific* (and *Male+Humanistic*).

A broad strand of literature in social psychology and an increasingly 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). However, 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)[31]. IAT scores are correlated with relevant behavior of individuals. 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).

Categories in the IAT can represent any grouping and it has been used to measure other form of implicit bias behind gender, as for instance race bias. Also in the context of race implicit bias,

[30]As in the standard IAT with a seven-block structure, individuals are asked to categorize in the first block only female and male words, in the second and fifth only scientific and humanistic subjects, while blocks three/ four and six/seven are those described in details and used for evaluation. Detailed explanation is provided in Appendix B.

[31]In particular, it has been shown that race bias (as measured by IAT) decreases after subjects viewed picture of admired African Americans and disliked White Americans (Dasgupta and Greenwald, 2001) or that stereotypes of white students are affected by friendship with black students (Burns et al., 2016).

studies have shown that IAT scores are correlated with call-back rates of minority job applicants (Rooth, 2010) and physicians' prescription of differential medical treatment by race (Green et al., 2007). Interestingly, Glover et al. (2017) provide evidence that interacting with bias individuals has negative self-fulfilling effects: they show that race bias of managers in French grocery stores negatively affect minority performance on the job. Despite substantial differences in terms of both type of bias (race vs. gender) and context (occupation vs. education), our paper relates to Glover et al. (2017) since we directly test the impact of implicit stereotypes of individuals on the stigmatized group they interact with.

Our measure of bias is strongly related to the context of our study: school performance and choice. Coherently with the purpose of our paper, we obtain a measure of teachers' gender bias in scholastic subjects (scientific vs. humanistic), interviewing teachers directly inside the school building. Hence, the measure of discrimination may suffer less to generalization of the bias to other contexts[32]. Furthermore, individuals complete the survey in the presence of a enumerator and therefore we are sure the teacher himself/herself completed the survey[33].

We collect teachers' implicit bias measure between the end of 2016 and the beginning of 2017 and we match these data with information from their past students who graduate from middle school between 2013 and 2015: the precise timeline is presented in Figure 1. Taking the IAT or knowledge about this study could not have affected neither students' performance nor teachers' or parents' attention to the issue of gender stereotypes for those cohorts of boys and girls. Teachers' experience in school may have shaped their implicit bias over the years. However, math teachers exploited 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, before and after the three cohort we are analyzing[34]. We will provide evidence that results are stable for the three cohort we analyze. Furthermore, the IAT score is calculated on field choice using subjects that are generally taught in the scientific academic track vs. humanistic tracks and we find no evidence of reverse causality in the choice of the

[32]Assume we were interested in evaluating the bias toward obese people in the work environment and we 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 in the job environment. 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 we want to evaluate.

[33]A less-expensive and time-consuming alternative could have been sending the survey by email. However, we were worried about both low response rate and identity of the individual completing the survey. Indeed, in the latter case, if she/he was not good with computers, she/he may have asked a children or partner to complete the survey on his/her behalf.

[34]For instance, students who were enrolled in middle school in the scholastic year 2015-2016 and 2016-2017 are not included in the sample. Usually, math teachers teach to 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 more than 80 students after the last cohort of students we analyze and before taking the IAT.

scientific track[35]. Finally, as it will be detailed in section 4.5, teachers with higher bias did not receive a different "treatment" in the type of students they were assigned to.

4.2 Administrative Data

Thanks to the authorization of each schools' principals, we were allowed to obtain and merge information from the Italian Ministry of Education and from the National Center for the Evaluation of the Italian Educational System (INVALSI). We collected individual level data for three cohort of students enrolled in grade 6 between the scholastic year 2010-11 and 2012-13 [36]. We have information about their math and reading standardized test score in grade 6 and 8 from INVALSI, together with information from their parents' education and occupation. We can match data from INVALSI with data from the Italian Ministry of Education about baseline individual information (date and place of birth, gender, citizenship) and the grades given by teachers at the end of each scholastic year. High-school track choice at the end of grade 8 and official teachers' recommendation is also provided. The timeline of data available for students is presented in Figure 1.

4.3 Student Survey

Administrative data on students can be merged with information collected by Carlana et al. (2017) through a survey to a sample of children in 24 schools among those involved in this research project[37]. We ask students to mention all subjects they will learn during high-school and to report their self-belief about own ability in each subject [38]. In all high-school, mathematics is taught and therefore most students report their self-assessment of math ability.

[35]In particular, we find that implicit bias of math teachers has no impact on scientific track choice and a negative impact on female students in the bottom of the distribution that are far from the margin of choosing a scientific track at high-school. The current bias of the teachers is not affects on average by the past choice toward a scientific high-school of their students.

[36]We have information on later cohorts as well, but it is not complete since the test score in grade 6 was not administered after 2012

[37] Carlana et al. (2017) evaluates a randomized control trial involving ten immigrant children per school. In order to collect information on soft skills of the control group, data are collected from all students in a random sample of 47 control schools. Of these 47 control schools, for 24 schools we have complete information on standardized test score and teachers' bias. The specific question we are interested in this paper was collected only for the control group and therefore not used in the paper by Carlana et al. (2017).

[38]The potential choices to that answer were three: "good", "mediocre", "scarce"

4.4 Descriptive Statistics

4.4.1 Teachers

Table 1 reports descriptive statistics of math teachers' information[39]. Most of teachers are females (84%), they are on average 52 years old with 23 years of experience in teaching and 92% holds a full-time contract. The majority of math teachers are born in a city in the North of Italy where the school are located (65%), but a substantial share is born in the Center or South of Italy and then migrated to the North. More than half of the teachers have a mother with education level lower than high-school diploma and 74% of them have at least one child (with an average of 1.84 children per teacher). Teachers graduating from math, physics and engineering are 24%, while most of them graduated from biology, natural sciences and other stem subjects. In the last part of Table 1, we report the summary statistics of explicit bias questions described in section 4.1. The variation in the answers on the equality of access to labor marker of men and women and about innate gender difference is ability is low, potentially also due to social desirability bias. It may be difficult to obtain revealed bias due to 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.

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, we standardize the IAT score to have mean zero and variance one throughout the paper. Considering the thresholds typically used in the social psychological literature[40], 25% of teachers are slightly or moderately in favor of girls, 30% present little to no bias, 19% show slight bias against female and 26% show moderate to severe bias against female. The sample of 1164 Italians used by Nosek et al. (2009) that decided to take the IAT online[41] 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 C.1). Interestingly, the great majority of math teachers are females and this may have important

[39] These are only teachers matched with students' information and therefore used in this paper. The main reasons for not being able to merge students' with teachers' data are twofold: teachers may have started to teach a class in grade 8 in the school after Summer 2015 or we did not obtained the authorization from the school signed on time to submit the data request (12 schools). Table A.4 shows the balance table of the differences between the sample of teachers matched and not matched with students' data. As expected, teachers not matched are around 9 years younger, 41 percent less likely to have full-time contract, have 12 years less of experience in teaching and 13 percent less likely to have children. However, as it can be clearly seen also from 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).

[40] 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.

[41] They completed the IAT online in the *Implicit Project* website.

implication for the association of scientific subjects with gender. Table 2 shows the correlation between math teacher IAT score and their characteristics. Females teahing math are strongly less bias in associating own gender with STEM and this aspect explain a substantial portion of the low average IAT score for math compared to Italian teachers (and also the sample used by Nosek et al. (2009)). Figure 2 plots the entire distributions 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 according with the IAT score. Teachers are more likely to associate own gender with the subject they teach, coherently 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.

In column 2 of Table 2, we correlate the IAT score with family and background characteristics: individuals born in the North are less gender bias than those born in the Center-South of Italy. This empirical evidence seems coherent with the belief that gender norms are stronger in the Center-South compared to the North of Italy, as suggested by differences in employment rate of women[42]. The point estimates of other family characteristics, as mother education and children gender, are generally small and statistically indistinguishable from zero. In column 3, we correlate the IAT score with qualifications of the teacher (type of degree, degree with honor), type of contract, other characteristics of the job and race IAT score. Individuals potentially with higher qualifications for teaching math, as for instance degree with honor and graduation from math, engineering and physics tend to have lower gender bias. We find no statistically significant correlation (and if anything a negative correlation) among the two conceptually unrelated measures of bias. Hence, the IAT score is not capturing a general "ability" in performing the test. Furthermore, in column 5, we correlate implicit and reported bias and we 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). Finally, in the last two column we correlate all teachers characteristics, respectively without and with school fixed effects, and we find that the patterns described above are confirmed. The same correlation are presented for the whole sample of teacher that completed our survey (not only those matched with students) in Table A.5 and the results are similar in terms of magnitude and significance.

[42]According to Istat data, in 2016 the employment rate of women was 31 percent in the South and 58 percent in the North.

4.4.2 Students

Table 3 reports balance tables of students' information obtain from administrative data and from the questionnaire with original first-hand data[43]. 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, we 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 to choose an academic scientific track and almost 25 percentage points less likely to enroll in a technical technological track. Females are more likely to choose an academic track than male, 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 equally chosen 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[44]. Finally, from the original first hand information available for a sample of students, we observe that on average there are no gender differences in assessment of ability, but females are 9 percentage points less likely to consider themselves good at math and boys are 5 percentage points less likely to consider themselves good in Italian.

4.5 Exogeneity Assumption

Before turning to analyze the empirical strategy, we present evidence of (1) as good as random assignment of students to classes in our sample of schools and that (2) gender biased teachers are not systematically assigned to specific groups of students (as for instance, classes with a higher share of females with low level of parents' education or lower ability in grade 6).
Within schools, classes are formed by the principal with the main objective of creating homogeneous groups in terms of gender, ability and socio-economic background and therefore to guarantee heterogeneity within one class[45]. We check whether students baseline characteristics

[43] We restrict the sample to students with information available on the standardized test score in grade 6 and 8 and for which we have the implicit association test of their math teacher in grade 6. This is exactly the sample that will be used in the empirical analysis of this paper. However, robustness checks will be provided considering attrition of students.

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

[45] As clarified in Section 3, among 81 principals, 64 percent consider the heterogeneity across classes in the ability level as "Extremely important" and 33 percent as "Important". The heterogeneity across classes in the socio-economic status is considered as "Extremely important" by 60 percent of principals and "Important" by 29

(gender, education and occupation of parents, immigration status and generation of immigration) and the class assignment are statistically independent with a series of Pearson Chi-Square tests. We find that around 6% of the tests performed, the p-value is lower or equal than 5%. This implies that for only 5% of the classes we cannot reject that there is non-random assignment in one of the baseline characteristics. We also perform the test recommended by Guryan et al. (2009) of random assignment of peers within school and cohort[46].

In Table 4, we also provide evidence that the characteristics of students are not systematically correlated with the implicit bias of teachers. We 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. We 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, we see that the difference is not statistically significant and the point estimate goes in the opposite direction. In columns 3, 4, 5 and 6, we analyze the correlation respectively with father occupation, immigration background and for the proxy of ability using standardized test scores in reading in grade 6[47] and we do not find statistically significant correlation and also very small point estimates. Finally, in the last column, we also include the standardized test score in math in grade 5, before entering middle school, despite the sample size is substantially reduced. The assumption of quasi-random assignment of students to teachers with different level of gender bias, as measured by the Implicit Association Test, within a school, is supported in the context we are analyzing.

5 The Impact of Teachers' Implicit Bias

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

$$y_{icts} = \alpha_0 + \alpha_1(Female_i \times bias_{cts}) + \alpha_2 Female_i + \eta_c + \mathbf{X}_{icts}\rho_1 + \varepsilon_{icts} \tag{3}$$

percent. The equal allocation of immigrant across classes is considered "Extremely Important" by 25 percent of principals and as "Important" by 38 percent.

[46] Guryan et al. (2009) proposes a well behaved test of random assignment, in which the characteristics of student i in class c of school s is regressed on the leave-one-out variable of own classmates and of own schoolmates in the same cohort, controlling for cohort by school fixed effects.

[47] We required standardized test score in math in grade 5, before students are assigned to teachers, but unfortunately we have obtained them only for those students that did not change school complex between elementary and middle school. Indeed, we have the authorization to access data only from the school principal of middle schools. Unfortunately, there are only few students per class for which we have this information.

where y_{icts} is an outcome of student i in class c taught by teacher t of school cohort s in grade 8 (as math standardized test score). $Female_i$ is a dummy variable which assumes value 1 if the student i is a girl and $bias_{cts}$ is the standardized value of the gender implicit bias of the professor teaching in class c in grade 8[48]. We include fixed effects at class level η_c, which absorb the average effect of the bias in class c, and controls for student characteristics $\mathbf{X}_{icts}$ (standardized test score in grade 6, parents education and occupation, immigration status and generation of immigration). Additionally, for robustness, the regression controls for student gender interacted with student characteristics $\mathbf{X}_{icts}$ and teacher characteristics $\mathbf{Z}_{cts}$ (as gender, place of birth, type of contract, type of degree, self-reported gender bias, ...)[49].

In Table 2 we have observed that IAT scores are correlated with some deeply-held characteristics (as gender and place of birth). All characteristics of teachers are absorbed by class fixed effect, but we additionally control for the interaction between these characteristics and the gender of the student in order to establish whether the impact of teacher stereotypes on gender gap among classmates can be explained (or attenuated) by teachers' observables. Standard errors are robust and clustered at teacher level.

The coefficient of interest, α_1, measures how the gender gap in the class changes when quasi-randomly assigned to teachers with one standard deviation higher bias. We expect the estimate of α_1 to be attenuated for the measurement error in the gender IAT score. The test-retest reliability of IAT is generally considered as satisfactory by social psychology (Nosek et al., 2007), with a correlation of 0.56 that does not change with the length of time between testing (despite being usually of less than one month in most studies). The IAT is expected to be the combination of the true implicit bias (a trait stable across time) and noise (occasion-specific variation)[50].

The second identification strategy relies on the comparison of students of the same gender enrolled in the same school, but assigned to different teachers. We investigate whether the impact of teacher bias on gender gap is due to reaction only of boys, girls or both. We estimate the equation

$$y_{icts} = \beta_0 + \beta_1(Female_i \times bias_{cts}) + \beta_2 Female_i + \beta_3 bias_{cts} + \eta_s + \mathbf{X}_{icts}\gamma_1 + \mathbf{Z}_{icts}\gamma_3 + \varepsilon_{icts} \quad (4)$$

where η_s are school cohort fixed effects and standard errors are clustered at teacher level. All other variables are defined as in equation (3). The coefficient of interest to evaluate the impact of implicit bias on male students is β_3, while the impact of females is obtained by $(\beta_1 + \beta_3)$.

[48]On average in 70% of cases teachers have been teaching to the same class from grade 6, in 11% of the cases from grade 7 and in 19% only for grade 8.

[49]The regression including all controls is:
$y_{icts} = \alpha_0 + \alpha_1(female_i \times bias_{cts}) + \alpha_2 female_i + \eta_c + \mathbf{X}_{icts}\rho_1 + (female_i \times \mathbf{X}_{icts})\rho_2 + (female_i \times \mathbf{Z}_{cts})\rho_4 + \varepsilon_{icts}$

[50]As suggested by Glover et al. (2017), we may expect an attenuation bias of approximately a factor of 1.8 due to measurement error in the IAT score.

As in the previous identification strategy, we will also control for the interaction between the gender of the student, own characteristics $\mathbf{X}_{icts}$ and teacher characteristics $\mathbf{Z}_{icts}$.

5.1 Performance in math

Table 5 shows the effect of teacher bias on gender gap in math performance within the class, presenting the results of estimating equation (3). The first two columns estimate the average gender gap in math performance: at the end of middle school females lag behind of around 0.22 standard deviation in the math standardized test score compared to their male classmates[51]. The additional gender gap in math performance between classmates from grade 6 to grade 8 is 0.078 standard deviations. The third column show that interacting with a teacher with one standard deviation higher bias leads to 0.027 standard deviation higher gender gap in math performance in the class, which corresponds to an increase of 34% of the gender difference in performance created during middle school. Adding the student controls and their interaction with the female dummy does not change the coefficient of interest as shown in column 4. In the Table, the coefficient of the interaction between math standardized test score in grade 6 and the gender of the student are always included but indistinguishable from zero.
The subsequent two columns add the interaction between the gender of the student and math teacher characteristics. If anything, the coefficient of interest *Fem*Bias Teacher*, which corresponds to α_1 in equation 3, slightly increases in magnitude. Observable characteristics of teachers, interacted with students' gender, are not driving the relation between gender gap and teacher bias. The results are robust to potential confounding aspects, even considering all information available on professors from their family background to their professional career. We report the coefficient only for the main characteristics of teachers interacted with students' gender for space reasons, but the effects are mainly small and insignificant for all variables. Ceteris paribus, female students assigned to female teachers or to teachers with an advanced STEM degree have lower math achievement test scores in grade 8 [52]. Having a teacher born in the North of the country does not have an heterogeneous effect on boys and girls. Finally, in column 6, we consider the impact of self-reported gender bias. Despite we acknowledge the possibility of social desirability bias in the response, it is interesting to notice that explicit bias has a negative impact on the gender gap in math performance, but this effect is on top of the impact of implicit bias. This evidence seems to support the distinctiveness of implicit and explicit cognition (Greenwald et al., 1998) in the context of gender stereotypes of math teacher and considering

[51] As shown by Table 3, the average gender gap without controlling for class fixed effects if anything is slightly smaller (0.21sd): most of the variation in math performance is not coming across schools or classes, but within, coherently with the target in class formation of heterogeneity within groups and homogeneity across groups.

[52] The impact of teacher gender is coherent with the result of Bharadwaj et al. (2016). Other papers find that having a teacher of own gender helps improving performance, especially at college level (Dee, 2005; Carrell et al., 2010).

our specific measure of self-reported bias that is strongly related to the most relevant reasons of gender differences in scientific track choice according with teachers.

We next investigate the effect of teacher bias from estimating equation (4), comparing students of the same gender withing the same school assigned to different classes. Table 6 presents the results 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 professor level included in columns 5 and 6 are important for the reliability of our results, but interestingly their inclusion do not significantly impact our main results. The self-reported gender bias has an effect only on female students, as found also by Alan et al. (2017) in the Turkish context, despite a different measure of self-reported bias of teachers.

5.1.1 Heterogeneous effects

We now examine the heterogeneous effect of teacher bias, first considering student characteristics and then interaction time with teachers. Table 7 shows that the effect is stronger for the most disadvantaged groups of students. Indeed, in column 1, we report the effect including teacher and student controls as in column 5 of Table 5. Column 2 investigates the effect of interest according with mother education and shows that one standard deviation higher bias of the teacher leads to an increase of the gender gap of 0.048 standard deviations among students with low educated mothers and of 0.025 standard deviations among students with higher level of education (at least a diploma), despite the difference is indistinguishable from zero. In column 3, we 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 (-0.068, with sd 0.026) and turning positive, but not statistically significant for students in the top of the initial ability distribution in grade 6. Finally, the effect seems stronger among immigrants (even if the difference with native is not statistically significant). Female students from less disadvantaged situation, as for instance 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 because for instance they have additional help or role models. Hence, this fact may explain the differential impact of teacher gender stereotypes.

The last two columns analyze whether there are heterogeneous effects according with the interaction time between students and teachers. Indeed, around 75% of students interact with the math teacher for six hours per week, while the rest of 9 hours per week. Furthermore, we exploit the fact that around 20% did not have the same teacher for all three years of middle school. However, for both variables, we do not see a statistically significant pattern.

5.2 Mechanism: Self-Stereotypes

One potential mechanism that can explain the impact of teacher bias on math performance of female students is self-stereotyping: girls' expectations about their group's suitability for math affect their achievement (Coffman, 2014). Biased math teacher may activate negative stereotypes and induce females to believe that they are "bad at math". Consistently with the concept of stereotype threat developed by social psychological literature (Steele and Aronson, 1995), teacher bias may stimulate self-stereotyping and induce individuals at risk of confirming widely-known negative stereotypes to underperform in fields in which their group is ability-stigmatized. In our conceptual framework in Section 2, we clarify the role of this particular mechanism. To test whether teacher activate self-stereotypes, we asked to a sample to students in grade 8 the extent to which they feel comfortable in their ability in different subjects. However, as we saw in section 5.1, gender gap is getting wider in those classes where teachers have higher bias. Hence, we also control for the mediating role of performance measured at the end of middle school in order to analyze whether there is an additional impact on self-stereotyping.

In Table 8, we assess the extent to which bias of teachers affect own assessment of ability, for a sample of around 800 students for which we collected these information (as described in section 4.3). As shown in column 1 respectively of Panel A and B, 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 both math and reading is increasing. In particular, in classes assigned to teachers with one standard deviation higher bias, the gender gap in self-assessment is increased by 4.5 percentage points. Adding student and teacher level controls do not substantially affect the point estimate of interest. In Section 5.1, we saw that the gender gap in math achievement increases in classes with higher bias. Hence, in the last three columns of Table 8, we examine whether gender gap in own assessment is due to the fact that gender gap in performance is truly bigger at the end of middle school. We find that indeed this is the case: the coefficient of interest of "Fem*Bias Teacher" decreases when controlling for the mediating factor of ability measured by standardized test scores. However, the point estimate is still negative and suggest that there is an additional effects of self-stereotypes, despite the small sample size leads to imprecise estimates. In the Appendix Table A.9, we show the result of the specification described in equation (4). The point estimates are imprecise, especially in the specification without controls, but the results points toward a negative impact on females and a slightly positive impact on males (even if not statistically significant in most specifications).

In Panel B and C, we focus on the impact on math teacher bias on self-assessment in other subjects different from math. More precisely, in Panel B, we focus on Italian, the other main subject taught during middle school in terms of number of hours, and in Panel C, we focus on the average on all other subjects. Students seems to compensate the low self assessment in math with higher self-assessment in Italian, but 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). Finally, in the last three columns of Panel B, we 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 (for further in-depth analysis of the impact on reading test scores or of the gender gap of the Italian teacher, we refer to Appendix C).

5.3 Additional Outcomes and Robustness Checks

5.3.1 Choice of High-School Track and Teachers Recommendation

High-school track choice is the first crucial career decision in the Italian schooling system. Teachers give an official track recommendation to their pupils, but students and their families are free to choose the track they like the most, with no constraints on grades and teachers' suggestions. As shown in Table 3, 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. Students in different tracks have in most of the cases little to no interaction during the school day since buildings and infrastructures are generally separated. Family background has been shown to play a crucial role in affecting track choice (Checchi et al., 2013), which is strongly correlated with university choice: 80% of graduates in STEM universities in 2015 did a scientific academic or a technical track during high-school (62% did the scientific academic high-school track).

We explore the impact of teacher bias on the track choice at the end of middle school. From a policy perspective the scientific academic path is particularly interesting since it easily opens up career opportunities in STEM related fields. However, in Table 9, we find a close to zero and insignificant effect of teacher bias on gender gap in scientific track choice (Panel A, columns 1-4) and in the recommendation of teachers toward a scientific track (Panel B, columns 1-4). In the questionnaire administered to teachers, we ask them why girls, compared to boys with the same math performance, are less likely to attend the scientific track: the reason considered as the most important is the influence of parents toward different tracks (for the summary statistics see Table 1). Furthermore, the scientific track is most likely chosen by females with highly educated parents and with high achievement tests, whose performance was not affected by teacher bias. These female students are likely to have additional academic-oriented role models on top of their math teacher and a lower vulnerability to the gender stereotypes.

Since our previous analysis showed that teacher stereotypes have stronger impact at the bottom of the ability distribution, we analyze also the choice of vocational track in Table 9, columns 5-8. Indeed, we observe that females, when assigned to a teacher with one standard deviation higher implicit bias, are more likely than their classmates to attend vocational track of around 2 percentage points. This effect mirrors an analogous differential in teachers' track recommendation toward vocational school as shown by Panel B, columns 6-8. Students move from technical technological track to vocational, as shown in the Appendix Table A.11.

5.3.2 Robustness checks

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.13, we check whether math teachers bias has an impact on retention rate, but we do not find any significant impact. Furthermore, we also check that teacher implicit stereotypes does not differentially impact the probability of attending the standardized test score in grade 8 (Table A.13, columns 5-8), conditional on attending 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.

Table 5 exploits information on three cohorts of students. In the appendix Table A.7, we show the effect for the three different cohort of students available: reassuringly, results are not statistically different in the three cohorts, even if since the number of observation decreases splitting the sample, estimates are noisier[53].

Finally, in the Table A.8, we present estimates of the impact on math performance of the Italian teacher bias, presenting the results of estimating equation (3). The gender bias of Italian teacher does not affect the gender gap in math performance. The point estimates is small and indistinguishable from zero. In the Appendix C, we delve deeper into the impact of Italian teacher on reading performance. Our results show that biased teachers activate stereotypes on female students only in male-typed domains.The differential response by gender and type of task is consistent with the previous results in the economic literature: subjects are negatively affected in gender incongruent areas, but the effect is particularly strong for females in male-typed domains. For instance, 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, both the environment (e.g. the sex composition

[53]For the first cohort, we have less observations because some schools change the code identifying the school that year for administrative reasons, but we are not allowed to access data identified with the older code.

of the setting) and the type of task has an effect on the willingness to complete Gneezy et al. (2003); Niederle and Vesterlund (2010). In particular, Große and Riener (2010) finds gender difference in competitiveness in stereotypically male tasks and no difference in stereotypically female tasks.

6 Conclusion

The cultural environment a person lives in has a strong impact on development of skills and educational choices. Indeed, we show that gender gap in math performance can be partially explained by teacher implicit bias. Females, especially those from disadvantaged backgrounds in terms of mother education and initial level of achievement, 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. Stereotypes foster low expectations about own ability and therefore induce a drop in performance. There are several possible mechanisms that may be involved simultaneously, as for instance anxiety, less (or less positive) interaction with the teacher, individual expectations about assessment of own ability. I provide evidence that the latter mechanism is a potentially important mediating factor. Our results show that biased teacher activate negative self-stereotypes on female students only in male-typed domains. Indeed, females are more likely to consider themselves bad in math at the end of middle school if they have a biased teacher, even controlling for their ability measured by standardized test scores. The findings are consistent with a model of stereotype whereby ability-stigmatized groups under-assess own ability and under-perform fulfilling negative expectations about their achievements. Teacher bias has also an impact on high-school track choice, leading female student assigned to a professor with higher stereotypes to be more likely to attend a vocational school. Unconscious biases and implicit associations can form an unintended and often an invisible barrier to equal opportunity.

These results raise the question on 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 0.078sd to 0.051sd). 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 in 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 stereotypes is provide alternative role models or higher confidence on own skills, 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 investigate further the impact of both type of policies. Nollenberger et al. (2016)

References

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

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

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

Arrow, K. J. et al. (1973). The theory of discrimination.

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.

Becker, G. S. et al. (1957). *Economics of Discrimination*. University of Chicago Press.

Bertrand, M., D. Chugh, and S. Mullainathan (2005). Implicit discrimination. *American Economic Review*, 94–98.

Bertrand, M. and E. Duflo (2016). Field experiments on discrimination. *NBER Working Paper*.

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. *Quarterly Journal of Economics*.

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

Burns, J., L. Corno, and E. La Ferrara (2016). Interaction, stereotypes and performance. evidence from south africa. *Working Paper*.

Carlana, M. (2017). Gender bias in grading. *Working Paper*.

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.

Checchi, D., C. V. Fiorio, and M. Leonardi (2013). Intergenerational persistence of educational attainment in italy. *Economics Letters 118*(1), 229–232.

Coate, S. and G. C. Loury (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 1220–1240.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, qju023.

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

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.

Dickerson, A., S. McIntosh, and C. Valente (2015). Do the maths: An analysis of the gender gap in mathematics in africa. *Economics of Education Review 46*, 1–22.

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. Technical report, Quarterly Journal of Economics.

Gneezy, U., M. Niederle, A. Rustichini, et al. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Green, A. R., D. R. Carney, D. J. Pallin, L. H. Ngo, K. L. Raymond, L. I. Iezzoni, and M. R. Banaji (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine 22*(9), 1231–1238.

Greenwald, A. G. and M. R. Banaji (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review 102*(1), 4.

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.

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.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

Hyde, J. S., E. Fennema, and S. J. Lamon (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin 107*(2), 139.

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.

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

Lundberg, S. J. and R. Startz (1983). Private discrimination and social intervention in competitive labor market. *The American Economic Review 73*(3), 340–347.

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.

Nollenberger, N., N. Rodríguez-Planas, and A. Sevilla (2016). The math gender gap: The role of culture. *The American Economic Review 106*(5), 257–261.

Nosek, B. A., M. R. Banaji, and A. G. Greenwald (2002). Math= male, me= female, therefore math$\neq$ me. *Journal of personality and social psychology 83*(1), 44.

Nosek, B. A., A. G. Greenwald, and M. R. Banaji (2007). The implicit association test at age 7: A methodological and conceptual review. *Automatic processes in social thinking and behavior*, 265–292.

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.

Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–61.

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.

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics 17*(3), 523–534.

Rudman, L. A., A. G. Greenwald, and D. E. McGhee (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin 27*(9), 1164–1178.

Spencer, S. J., C. M. Steele, and D. M. Quinn (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology 35*(1), 4–28.

Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology 69*(5), 797.

Terrier, C. (2015). Giving a little help to girls? evidence on grade discrimination and its effect on students' achievement.

Tversky, A. and D. Kahneman (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pp. 141–162. Springer.

Figures and Tables

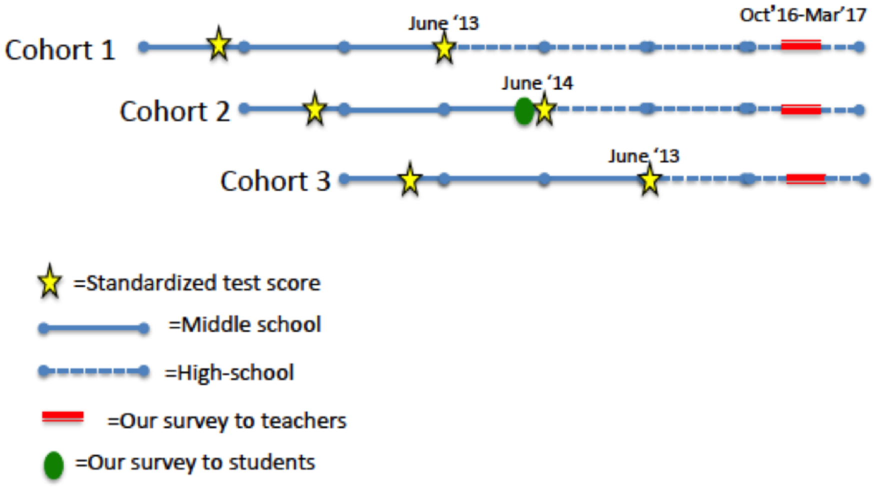

Figure 1: Timeline of main data available for students and teachers

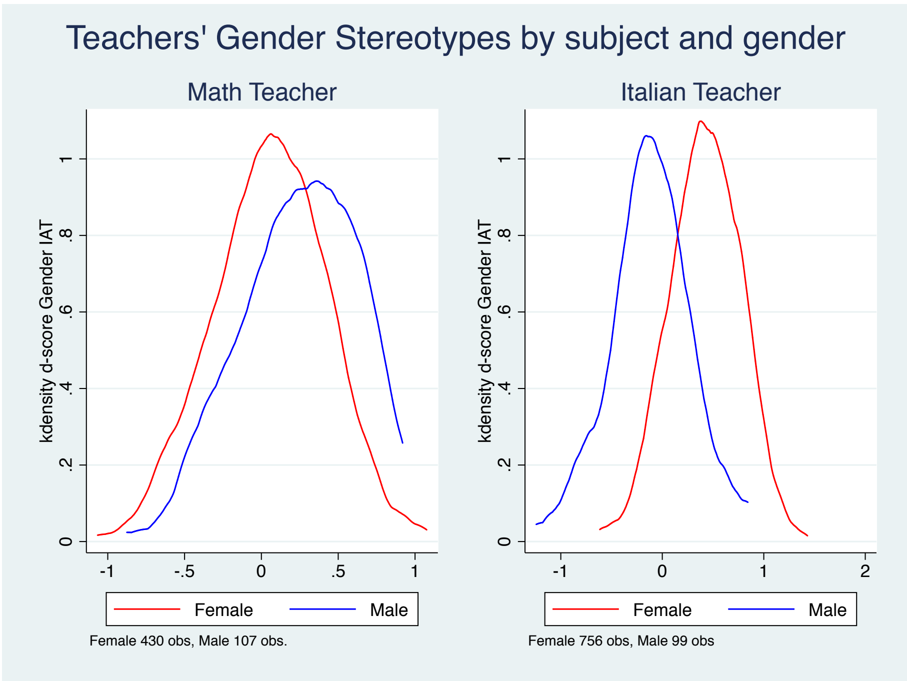

Figure 2: Teachers' Implicit Gender Bias (IAT measure) by gender and subject they teach

Table 1: Summary Statistics from Math Teachers' Questionnaire

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
IAT Race	301	0.46	0.27	-0.38	1.11
Self-reported explicit bias					
WVS Gender Equality	285	1.14	0.36	0.00	2.00
Gender Dif Innate Ability	280	1.56	0.87	1.00	5.00
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
Reported gender bias	206	0.00	1.01	-1.41	1.91
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
hline 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.4. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

Table 2: Correlation between math teachers' characteristics and Gender IAT Score

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: raw IAT score of math teachers						
Female	-0.168***	-0.167***			-0.174***	-0.184**
	(0.057)	(0.056)			(0.056)	(0.085)
Born in the North		-0.067			-0.056	-0.127*
		(0.044)			(0.048)	(0.075)
Age		0.048			0.034	0.036
		(0.039)			(0.038)	(0.063)
Age sq.		-0.000			-0.000	-0.000
		(0.000)			(0.000)	(0.001)
Children		-0.015			-0.011	0.041
		(0.054)			(0.056)	(0.075)
Daughters		0.103			0.104	0.141
		(0.096)			(0.092)	(0.106)
Middle edu Mother		0.004			0.017	-0.001
		(0.049)			(0.048)	(0.066)
High edu Mother		-0.018			-0.030	0.003
		(0.060)			(0.065)	(0.090)
Advanced STEM			-0.061		-0.068	-0.098
			(0.051)		(0.052)	(0.074)
Degree Laude			-0.109*		-0.063	-0.032
			(0.055)		(0.057)	(0.085)
Math Olympiad			0.071		0.109*	0.081
			(0.062)		(0.065)	(0.089)
Full time contract			-0.021		-0.057	-0.071
			(0.094)		(0.095)	(0.141)
Satisfy with teacher job			0.040*		0.035	0.040
			(0.024)		(0.024)	(0.035)
IAT Race			-0.137		-0.143	-0.087
			(0.096)		(0.095)	(0.117)
Explicit Bias				0.030	0.012	0.027
				(0.024)	(0.025)	(0.039)
Constant	0.267***	-0.997	0.080	0.127***	-0.709	-0.759
	(0.055)	(0.957)	(0.127)	(0.034)	(0.949)	(1.493)
School FE	No	No	No	No	No	Yes
Obs.	301	301	301	301	301	301
R^2	0.040	0.075	0.058	0.016	0.129	0.426

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. 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, "Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. We include the education of the mother of the teacher, the type of degree and whether is was achieved with laude, the type of contract and other administrative responsibilities within the school. Check Table A.5 for the same correlation for the sample of all math teachers available. 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 3: Summary Statistics of students by gender

	Males	Females	Diff.	se
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)
Average own ability	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)
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)
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 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)	(6)
Fem	-0.222***	-0.078***	-0.080***	-0.036	-0.012	-0.054
	(0.019)	(0.014)	(0.014)	(0.032)	(0.104)	(0.106)
Fem*Bias Teacher			-0.027**	-0.028**	-0.036***	-0.034**
			(0.013)	(0.013)	(0.014)	(0.013)
Fem*Teacher Fem					-0.058	-0.049
					(0.036)	(0.036)
Fem*Born North Teacher					0.013	0.015
					(0.030)	(0.029)
Fem*Advanced STEM Teacher					-0.046	-0.041
					(0.031)	(0.030)
Fem*Reported Bias Teacher						-0.049***
						(0.016)
Std Math grade 6		0.723***	0.723***	0.697***	0.698***	0.698***
		(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Constant	0.198***	0.028***	0.028***	-0.112***	-0.113***	-0.110***
	(0.009)	(0.007)	(0.007)	(0.023)	(0.023)	(0.023)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	Yes
Teacher Controls	No	No	No	No	Yes	Yes
Obs.	9309	9309	9309	9309	9309	9309
R^2	0.209	0.618	0.618	0.625	0.625	0.626

Notes: This table reports OLS estimates of equation 3, 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, 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 *** 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)	(6)
Fem	-0.234***	-0.092***	-0.093***	-0.034	-0.019	-0.055
	(0.022)	(0.015)	(0.015)	(0.033)	(0.108)	(0.108)
Fem*Bias Teacher			-0.022*	-0.024*	-0.032**	-0.030**
			(0.013)	(0.013)	(0.013)	(0.013)
Bias Teacher			-0.011	-0.011	-0.006	-0.007
			(0.015)	(0.014)	(0.013)	(0.013)
Fem*Math Teacher Fem					-0.052	-0.045
					(0.039)	(0.039)
Math Teacher Fem					0.065	0.063
					(0.040)	(0.040)
Fem*North Math Teacher					0.018	0.019
					(0.030)	(0.029)
Math Teacher born North					0.024	0.022
					(0.034)	(0.034)
Fem*Advanced STEM Teacher					-0.033	-0.032
					(0.034)	(0.033)
Advanced STEM					0.028	0.028
					(0.034)	(0.034)
Fem*Reported Bias Teacher						-0.055***
						(0.016)
Reported Bias Teacher						0.017
						(0.016)
Std Math grade 6		0.716***	0.715***	0.687***	0.688***	0.687***
		(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Constant	0.190***	0.023	-0.001	-0.156***	-0.304**	-0.293**
	(0.020)	(0.015)	(0.019)	(0.028)	(0.126)	(0.126)
School, year FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	Yes
Teacher Controls	No	No	No	No	Yes	Yes
Obs.	9309	9309	9309	9309	9309	9309
R^2	0.136	0.576	0.577	0.585	0.588	0.588

Notes: This table reports OLS estimates of equation 4, 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, 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, self-reported gender bias 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	-0.012	-0.007	0.045	-0.011	-0.039	-0.020
	(0.104)	(0.104)	(0.113)	(0.104)	(0.104)	(0.104)
Fem*Bias Teacher	-0.036***	-0.048**	-0.068**	-0.034**	-0.039**	-0.062**
	(0.014)	(0.021)	(0.026)	(0.015)	(0.016)	(0.031)
Fem*Bias T*HighEduM		0.023				
		(0.028)				
Fem*Bias T*Top tercile Math6			0.101***			
			(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					0.016	
					(0.026)	
Fem*Bias T*Same Math Teacher						0.029
						(0.035)
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	0.625	0.626	0.627	0.625	0.625	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, 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. 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 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	0.312	-0.053*	-0.074	0.297
	(0.029)	(0.028)	(0.065)	(0.197)	(0.028)	(0.065)	(0.205)
Fem*Bias Teacher		-0.045**	-0.049**	-0.062**	-0.030	-0.033	-0.048
		(0.021)	(0.022)	(0.029)	(0.021)	(0.023)	(0.029)
Constant	0.837***	0.808***	0.809***	0.798***	0.810***	0.820***	0.810***
	(0.015)	(0.015)	(0.048)	(0.049)	(0.015)	(0.048)	(0.047)
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	0.250	0.248	0.266	0.278
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Female	0.052**	0.057**	0.045	0.288	0.047**	0.035	0.256
	(0.023)	(0.023)	(0.048)	(0.194)	(0.021)	(0.046)	(0.187)
Fem*Bias Teacher		0.038**	0.038**	0.025	0.038**	0.039**	0.027
		(0.018)	(0.019)	(0.019)	(0.017)	(0.019)	(0.018)
Constant	0.916***	0.908***	0.937***	0.950***	0.917***	0.953***	0.966***
	(0.012)	(0.012)	(0.034)	(0.034)	(0.011)	(0.034)	(0.034)
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	0.185	0.148	0.161	0.198
Panel C- Dependent Variable: Average own ability in other subjects							
Female	0.035	0.019	0.021	-0.241	0.016	0.018	-0.241
	(0.027)	(0.029)	(0.062)	(0.213)	(0.028)	(0.062)	(0.212)
Fem*Bias Teacher		-0.014	-0.015	-0.029	-0.018	-0.020	-0.033
		(0.023)	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)
Constant	1.672***	1.689***	1.674***	1.676***	1.687***	1.670***	1.672***
	(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	0.160	0.130	0.141	0.164
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 3, 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, 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 *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table 9: Estimation of the effet of teachers' gender stereotypes on track choice- class FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.094***	-0.048***	0.178*	0.174*	0.014	-0.009	0.017	0.020
	(0.012)	(0.011)	(0.092)	(0.092)	(0.009)	(0.010)	(0.070)	(0.071)
Fem*Bias Teacher		0.009	0.001	0.001		0.023**	0.022**	0.021**
		(0.012)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Fem*Reported Bias				-0.003				0.003
				(0.013)				(0.011)
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***	0.108***	0.108***	0.141***	0.174***	0.205***	0.205***
	(0.006)	(0.006)	(0.015)	(0.015)	(0.005)	(0.006)	(0.016)	(0.016)
Obs.	8463	8463	8463	8463	8463	8463	8463	8463
R^2	0.113	0.214	0.235	0.236	0.119	0.190	0.210	0.210
	Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.045***	-0.019**	0.022	0.013	-0.059***	-0.110***	-0.070	-0.081
	(0.010)	(0.009)	(0.082)	(0.080)	(0.013)	(0.011)	(0.091)	(0.089)
Fem*Bias Teacher		0.001	-0.007	-0.006		0.018*	0.024**	0.025**
		(0.009)	(0.009)	(0.009)		(0.010)	(0.011)	(0.011)
Fem*Reported Bias				-0.015				-0.010
				(0.012)				(0.012)
Std Math grade 6		0.126***	0.113***	0.113***		-0.246***	-0.218***	-0.218***
		(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***	0.517***	0.518***
	(0.005)	(0.004)	(0.011)	(0.011)	(0.006)	(0.006)	(0.017)	(0.017)
Obs.	7086	7086	7086	7086	7086	7086	7086	7086
R^2	0.152	0.238	0.251	0.251	0.150	0.362	0.390	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	Yes	Yes	No	No	Yes	Yes

Notes: This table reports OLS estimates of equation 3, 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, 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 *** indicate significance at the 10%, 5% and 1% percent level respectively.

A Appendix: Additional Tables and Figures

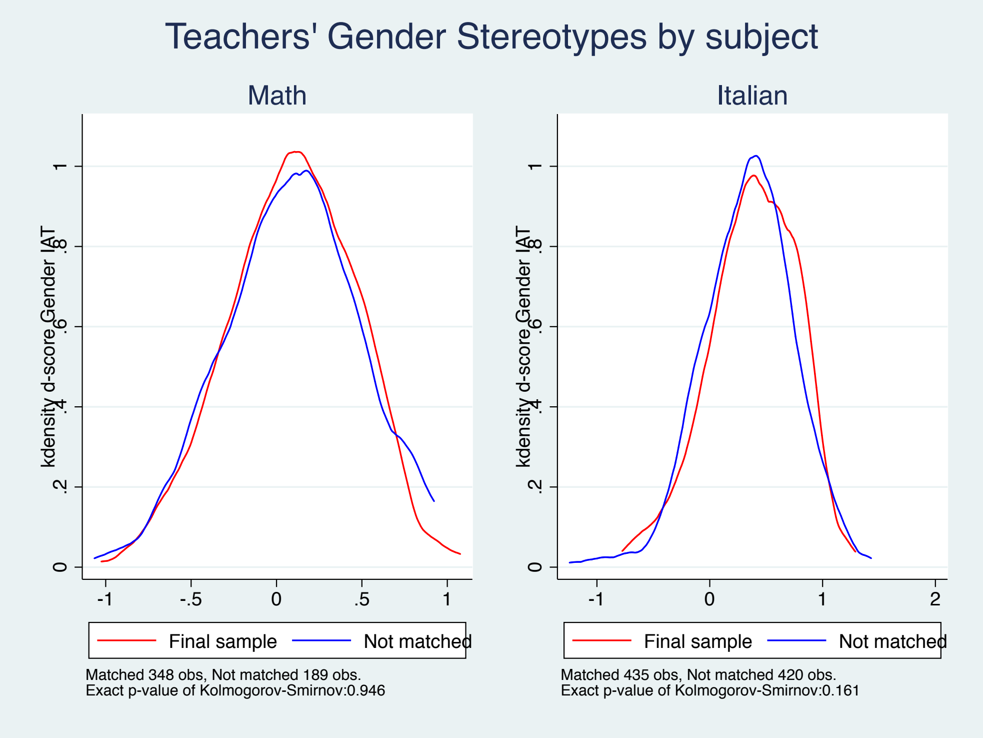

Figure A.1: Teachers' Implicit Gender Bias (IAT measure) by subject of matched and unmatched sample

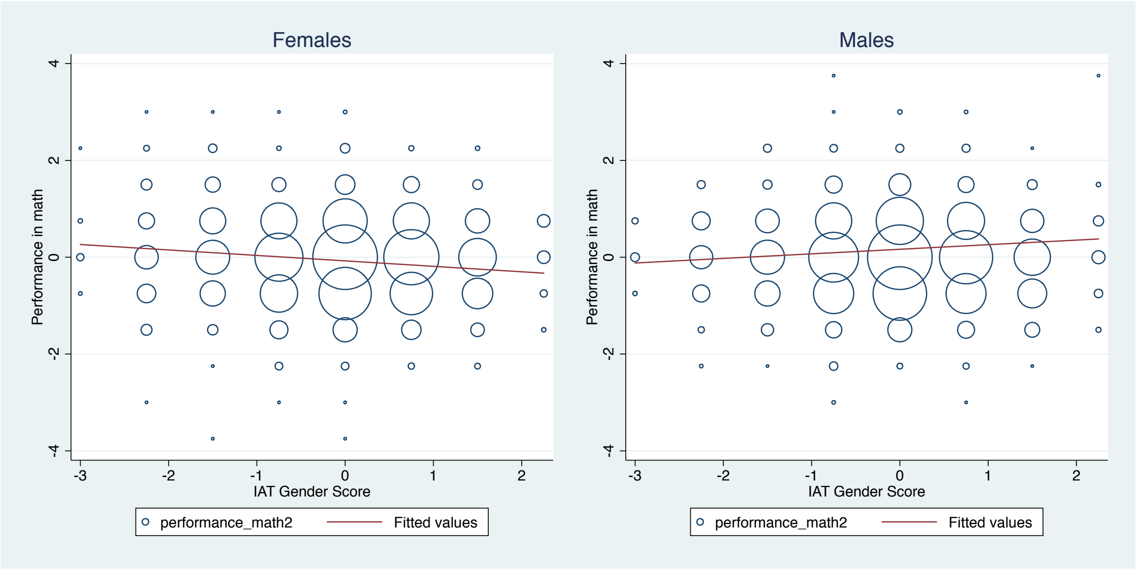

Figure A.2: Correlation between the performance in math and the implicit bias of math teachers (as measured by IAT)

Table A.1: Correlation between subject, gender and own assessment

Dependent Variable: Performing well because she/he is good at the subject (math/reading)

Gender of students	Males		Female		All			
Subject	Reading and Math		Reading and Math		Math	Reading	Reading and Math	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	-0.026***	-0.025***	-0.087***	-0.087***			-0.025***	-0.026***
	(0.006)	(0.006)	(0.005)	(0.005)			(0.006)	(0.008)
Math*Female							-0.062***	-0.062***
							(0.008)	(0.011)
Female					-0.076***	-0.023***	-0.018***	
					(0.005)	(0.006)	(0.006)	
Math test score		0.002***		0.001***	0.003***		0.001***	
		(0.000)		(0.000)	(0.000)		(0.000)	
Italian Test score		0.000		-0.000		0.000***	-0.000	
		(0.000)		(0.000)		(0.000)	(0.000)	
Constant	0.195***	0.098***	0.174***	0.148***	0.024***	0.164***	0.128***	0.185***
	(0.004)	(0.012)	(0.004)	(0.013)	(0.007)	(0.012)	(0.009)	(0.003)
Individual FE	No	No	No	No	No	No	No	Yes
Obs.	17537	17432	16491	16403	16952	16979	33835	34028
R^2	0.001	0.010	0.017	0.018	0.042	0.001	0.018	0.531

Robust Standard Errors clustered at individual level in parentheses. This information is collected together with the standardized test score in grade 6. "Female" refers to the gender of students. "Math" is a dummy which assumes value 1 if the dependent variable is related to own assessment in math and 0 if it is related to own assessment in reading.

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

Table A.2: Factors loadings: Reported gender bias

	Math Teachers	Italian Teachers
WVS Gender Equality	0.2537	0.0257
Gender Dif Innate Ability	0.2206	0.2372
Reason GenderGap: Cultural Stereotypes	0.2591	0.3284
Reason GenderGap: Interest for STEM	0.8163	0.7870
Reason GenderGap: Low self-esteem	0.6317	0.7338
Reason GenderGap: Predisposition for STEM	0.8235	0.7563

Factor analysis is performed using polychoric correlation matrix since variables are not continuous and do not follow a multivariate normal distribution (ordinal variables)

Table A.3: Correlation between implicit bias d-score and order of different parts of the survey

Dependent variable : d-score Implicit Bias Math Teachers	(1)	(2)	(3)	(4)	(5)
First IAT Gender	-0.035			-0.034	-0.040
	(0.037)			(0.037)	(0.047)
First Questionnaire, then IAT		-0.151		-0.166	0.017
		(0.127)		(0.126)	(0.185)
Order Compatible IAT Gender			-0.051*	-0.052*	-0.049
			(0.029)	(0.030)	(0.039)
Constant	0.106***	0.097***	0.119***	0.134***	0.131***
	(0.019)	(0.016)	(0.020)	(0.022)	(0.024)
Obs.	534	534	534	534	534
R^2	0.002	0.002	0.005	0.009	0.168

Notes: This table reports OLS estimates of the correlation between order of IAT and IAT score. A higher value of IAT score means stronger implicit association between Male-Science and Female-Literature. The dummy "First IAT Gender" captures the order of IATs (genderand race). The variable "Order Compatible IAT Gender" captures whether it was asked to associate together first more likely compatible categories (Male-Scientific vs. Female-Humanistic) or the opposite (Female-Scientific vs. Male-Humanistic). Finally, in 8 cases for the math teacher and 32 cases for the Italian teacher we asked to complete first a questionnaire and then the IATs. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.4: Balance table of the differences between teachers matched (not matched) with students graduating from 2013 to 2015

	Math Teachers				Italian Teachers			
	Not matched	Matched	Dif.	se	Not matched	Matched	Dif.	se
Female	0.763	0.837	-0.074*	(0.034)	0.858	0.910	-0.052*	(0.022)
Born in the North	0.545	0.653	-0.108*	(0.043)	0.606	0.737	-0.131***	(0.033)
Age	42.569	51.903	-9.335***	(0.767)	42.889	51.367	-8.477***	(0.594)
Full time contract	0.516	0.916	-0.400***	(0.035)	0.731	0.988	-0.257***	(0.022)
Yeas of experience	10.646	22.941	-12.295***	(0.938)	13.854	23.695	-9.842***	(0.691)
Teaching in 2015-16	0.560	0.987	-0.427***	(0.030)	0.617	1.000	-0.383***	(0.023)
Children	0.572	0.744	-0.172***	(0.040)	0.569	0.729	-0.160***	(0.032)
Number of children	1.858	1.842	0.016	(0.095)	1.674	1.786	-0.112	(0.074)
Number of daughters	0.890	0.851	0.039	(0.087)	0.916	0.815	0.101	(0.068)
Low edu Mother	0.415	0.579	-0.164***	(0.045)	0.387	0.516	-0.129***	(0.036)
Middle edu Mother	0.401	0.291	0.110*	(0.043)	0.403	0.376	0.027	(0.036)
High edu Mother	0.184	0.129	0.054	(0.033)	0.210	0.108	0.102***	(0.026)
Advanced STEM	0.237	0.240	-0.003	(0.038)	0.000	0.000	0.000	(0.000)
Math Olympiad	0.088	0.192	-0.104***	(0.031)	0.000	0.000	0.000	(0.000)
Update Courses	0.851	0.938	-0.087***	(0.026)	0.863	0.932	-0.069**	(0.021)
Degree Laude	0.276	0.168	0.108**	(0.039)	0.386	0.293	0.093**	(0.035)
IAT Gender	0.104	0.087	0.016	(0.033)	0.346	0.386	-0.040	(0.027)
IAT Race	0.472	0.458	0.014	(0.023)	0.450	0.464	-0.014	(0.018)
Boys better in Invalsi	0.241	0.202	0.039	(0.043)	0.095	0.099	-0.003	(0.026)
Girls better in Invalsi	0.304	0.322	-0.018	(0.048)	0.550	0.527	0.023	(0.043)
Gender Equal in Invalsi	0.456	0.476	-0.021	(0.052)	0.355	0.374	-0.020	(0.042)
Satisfy with teacher job	3.749	3.692	0.057	(0.079)	3.862	3.890	-0.028	(0.061)
WVS Gender Equality	1.108	1.137	-0.029	(0.032)	1.071	1.069	0.003	(0.021)
Reason GenderGap: Interest for STEM	2.543	2.579	-0.036	(0.095)	2.869	2.648	0.221**	(0.076)
Reason GenderGap: Predisposition for STEM	2.127	2.124	0.003	(0.103)	2.222	2.158	0.064	(0.081)
Reason GenderGap: Low self-esteem	2.905	2.638	0.267**	(0.094)	2.694	2.540	0.154	(0.079)
Reason GenderGap: Family support	3.155	3.144	0.011	(0.098)	3.118	2.941	0.176*	(0.076)
Reason GenderGap: Cultural Stereotypes	2.461	2.147	0.314**	(0.105)	2.316	2.166	0.150	(0.086)
Gender Dif Innate Ability	1.629	1.559	0.070	(0.081)	1.540	1.428	0.113	(0.060)
Reported gender bias	-0.003	0.002	-0.005	(0.106)	-0.061	0.059	-0.120	(0.083)
Observations	236	301			422	431		

Notes: First hand data from teachers' questionnaire. We compare teachers' matched with students' data with those not matched. The main reasons for not being able to merge students' with teachers' data are twofold: teachers may have started to teach a class in grade 8 in the school after Summer 2015 or we did not obtained the authorization from the school signed on time to submit the data request (12 schools).

Table A.5: Correlation between math teachers' characteristics and Gender IAT Score

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: raw IAT score of math teachers						
Female	-0.193***	-0.185***			-0.190***	-0.171***
	(0.043)	(0.043)			(0.044)	(0.059)
Born in the North		-0.077**			-0.068**	-0.115**
		(0.033)			(0.034)	(0.044)
Age		-0.011			-0.022	-0.039*
		(0.017)			(0.018)	(0.023)
Age sq.		0.000			0.000	0.000
		(0.000)			(0.000)	(0.000)
Children		0.029			0.029	0.039
		(0.040)			(0.040)	(0.048)
Daughters		0.041			0.044	0.075
		(0.068)			(0.066)	(0.076)
Middle edu Mother		0.021			0.020	0.007
		(0.040)			(0.040)	(0.049)
High edu Mother		-0.010			-0.004	0.006
		(0.043)			(0.046)	(0.057)
Advanced STEM			-0.084**		-0.093**	-0.142***
			(0.042)		(0.042)	(0.050)
Degree Laude			-0.050		-0.062	-0.051
			(0.039)		(0.042)	(0.052)
Math Olympiad			0.063		0.078	0.101
			(0.050)		(0.052)	(0.069)
Full time contract			-0.047		0.009	-0.005
			(0.046)		(0.055)	(0.073)
Satisfy with teacher job			0.020		0.009	0.012
			(0.016)		(0.017)	(0.020)
IAT Race			-0.074		-0.092	-0.078
			(0.075)		(0.071)	(0.084)
Explicit Bias				0.026	0.017	0.033
				(0.020)	(0.020)	(0.026)
Constant	0.285***	0.574	0.150*	0.131***	0.870**	1.318**
	(0.041)	(0.402)	(0.085)	(0.027)	(0.431)	(0.539)
School FE	No	No	No	No	No	Yes
Obs.	534	534	533	534	533	533
R^2	0.050	0.065	0.039	0.012	0.097	0.288

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. 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, "Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. We include the education of the mother of the teacher, the type of degree and whether is was achieved with laude, the type of contract and other administrative responsibilities within the school. Check Table 2 for the same correlation for the sample of math teachers matched with complete student data. 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 A.6: 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)	(8)
Share Female	0.883					0.739	0.506	-1.610
	(1.013)					(1.051)	(1.006)	(3.136)
Share Male HighEduMother		-0.117				-0.246	-0.173	0.793
		(0.708)				(0.719)	(0.752)	(2.625)
Share Female HighEduMother		0.573				0.449	0.502	-0.310
		(0.606)				(0.595)	(0.665)	(2.137)
Share Male HighOccFather			0.402			0.861	1.078	0.888
			(1.152)			(1.217)	(1.076)	(3.833)
Share Female HighOccFather			0.289			-0.074	-0.074	0.589
			(0.744)			(0.897)	(0.972)	(2.258)
Share Male MedOccFather			0.222			0.750	0.636	0.478
			(0.764)			(0.738)	(0.780)	(1.251)
Share Female MedOccFather			1.055			0.869	0.500	0.173
			(0.816)			(0.881)	(0.887)	(1.985)
Share Male Immigrant				0.851		1.113*	1.162*	-0.122
				(0.638)		(0.637)	(0.623)	(1.977)
Share Female Immigrant				-0.218		-0.080	-0.102	0.560
				(0.619)		(0.669)	(0.701)	(1.255)
Male Average Std Ita6					-0.145	0.031	-0.055	-0.500
					(0.273)	(0.301)	(0.314)	(1.333)
Female Average Std Ita6					0.220	0.082	0.145	0.991
					(0.271)	(0.312)	(0.305)	(0.855)
Male Average Std Mat5								-0.187
								(0.593)
Female Average Std Mat5								-0.458
								(0.422)
Constant	-0.368	-0.112	-0.572	-0.081	0.030	-1.400	-1.529	-0.910
	(0.516)	(0.444)	(0.697)	(0.163)	(0.103)	(0.941)	(1.371)	(3.250)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	No	Yes	Yes
Obs.	301	301	301	301	301	301	301	110
R^2	0.329	0.332	0.341	0.337	0.328	0.366	0.444	0.590

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and students' characteristics; the unit of observation is student *i*, in class *c* taught by teacher *t* in grade 8 of school *s*. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90 in columns 1-6 and 40 in column 7. Teacher controls include teacher gender, place of birth, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias and their interactions with students' gender. All regression include controls for the order of IAT in the questionnaire administered. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.7: Estimation of the effect of teachers' gender stereotypes on standardized test score in math in grade 8 for the different cohorts

	Dependent Variable: Math standardized test score in grade 8			
	All Students (1)	First Cohort (2)	Second Cohort (3)	Third Cohort (4)
Female	-0.012	-0.041	-0.055	-0.064
	(0.104)	(0.274)	(0.169)	(0.174)
Fem*BiasTeacher	-0.036***	-0.041	-0.033	-0.038*
	(0.014)	(0.035)	(0.022)	(0.020)
Constant	-0.113***	0.050	-0.183***	-0.112***
	(0.023)	(0.053)	(0.037)	(0.031)
Class FE	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	No	Yes	Yes
Obs.	9309	1984	4143	3182
R^2	0.625	0.623	0.608	0.661

Notes: This table reports OLS estimates of equation 3, 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, 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 A.8: Estimation of the effect of Italian 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)
Fem	-0.055***	-0.043	0.553***	0.553***
	(0.016)	(0.033)	(0.140)	(0.137)
Fem*Bias Ita Teacher	0.005	0.002	0.006	0.006
	(0.016)	(0.016)	(0.017)	(0.017)
Fem*Ita Teacher Fem			-0.052	-0.050
			(0.067)	(0.068)
Fem*Born North Ita Teacher			-0.045	-0.047
			(0.037)	(0.038)
Fem*Reported Bias Ita Teacher				0.012
				(0.018)
Std Math grade 6	0.722***	0.695***	0.695***	0.695***
	(0.012)	(0.013)	(0.013)	(0.013)
Constant	0.016*	-0.105***	-0.105***	-0.105***
	(0.008)	(0.024)	(0.024)	(0.024)
Class FE	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes
Teacher Controls	No	No	Yes	Yes
Obs.	8644	8644	8644	8644
R^2	0.612	0.619	0.620	0.621

Notes: This table reports OLS estimates of equation 3, where the dependent variable is math standardized test score in grade 8; the unit of observation is student i, in class c taught by Italian teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 352. The number of fixed effects (classes) is 504. 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, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.9: Estimation of the effet of teachers' gender stereotypes on self-stereotypes- school FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A- Dependent Variable: Being good/mediocre at math (vs. being bad)							
Female	-0.124***	-0.092***	-0.102*	0.240	-0.073***	-0.075	0.231
	(0.028)	(0.027)	(0.061)	(0.187)	(0.027)	(0.061)	(0.191)
Fem*Bias Teacher		-0.035	-0.043*	-0.057*	-0.021	-0.028	-0.045
		(0.021)	(0.023)	(0.029)	(0.021)	(0.023)	(0.029)
Bias Teacher		0.009	0.016	0.032*	0.015	0.022	0.029*
		(0.016)	(0.016)	(0.018)	(0.016)	(0.016)	(0.017)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	783	783	783	783	783	783	783
R^2	0.073	0.191	0.211	0.244	0.223	0.240	0.275
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Female	0.046*	0.045*	0.037	0.296	0.037	0.028	0.269
	(0.025)	(0.025)	(0.049)	(0.187)	(0.023)	(0.047)	(0.181)
Fem*Bias Teacher		0.038**	0.040**	0.028	0.039**	0.041**	0.030*
		(0.017)	(0.017)	(0.018)	(0.016)	(0.016)	(0.017)
Bias Teacher		-0.005	-0.006	0.004	-0.003	-0.004	0.004
		(0.014)	(0.013)	(0.015)	(0.014)	(0.014)	(0.015)
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	0.043	0.068	0.085	0.151	0.080	0.092	0.160
Panel C- Dependent Variable: Average own ability in all other subjects							
Female	0.038	0.025	0.020	-0.275	0.022	0.015	-0.281
	(0.029)	(0.031)	(0.062)	(0.208)	(0.031)	(0.063)	(0.206)
Fem*Bias Teacher		-0.011	-0.011	-0.020	-0.013	-0.014	-0.023
		(0.025)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)
Bias Teacher		0.021	0.022	0.026*	0.018	0.019	0.024
		(0.019)	(0.019)	(0.015)	(0.019)	(0.019)	(0.016)
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	0.030	0.059	0.072	0.130	0.063	0.075	0.135
School, Cohort 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 4, 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 (school, cohort) is 23. 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, 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, self-reported gender bias and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.10: Estimation of the effet of teachers' gender stereotypes on track choice- school FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.093***	-0.049***	0.157*	0.151*	0.004	-0.018*	0.004	0.008
	(0.014)	(0.013)	(0.088)	(0.088)	(0.010)	(0.011)	(0.067)	(0.067)
Fem*Bias Teacher		0.008	0.002	0.003		0.024***	0.023***	0.023**
		(0.011)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Bias Teacher		-0.007	-0.005	-0.005		-0.014**	-0.011	-0.010
		(0.008)	(0.008)	(0.008)		(0.007)	(0.007)	(0.007)
Std Math grade 6		0.171***	0.152***	0.152***		-0.103***	-0.090***	-0.089***
		(0.007)	(0.007)	(0.007)		(0.006)	(0.006)	(0.006)
Constant	0.301***	0.248***	0.017	0.019	0.141***	0.172***	0.229***	0.227***
	(0.011)	(0.010)	(0.074)	(0.074)	(0.008)	(0.008)	(0.051)	(0.052)
Obs.	8463	8463	8463	8463	8463	8463	8463	8463
R^2	0.067	0.171	0.197	0.197	0.068	0.144	0.169	0.169
	Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.058***	-0.031***	0.004	-0.001	-0.058***	-0.110***	-0.093	-0.099
	(0.012)	(0.011)	(0.079)	(0.078)	(0.015)	(0.013)	(0.094)	(0.092)
Fem*Bias Teacher		0.000	-0.007	-0.007		0.018*	0.022**	0.023**
		(0.009)	(0.009)	(0.009)		(0.010)	(0.010)	(0.010)
Bias Teacher		-0.006	-0.002	-0.003		-0.015*	-0.015*	-0.016**
		(0.008)	(0.007)	(0.007)		(0.008)	(0.008)	(0.008)
Std Math grade 6		0.124***	0.111***	0.110***		-0.237***	-0.209***	-0.210***
		(0.009)	(0.008)	(0.008)		(0.008)	(0.008)	(0.008)
Constant	0.171***	0.147***	0.052	0.055	0.366***	0.412***	0.483***	0.485***
	(0.011)	(0.010)	(0.067)	(0.065)	(0.015)	(0.013)	(0.080)	(0.076)
Obs.	7086	7086	7086	7086	7086	7086	7086	7086
R^2	0.097	0.187	0.202	0.203	0.103	0.321	0.355	0.356
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	Yes	Yes	No	No	Yes	Yes

Notes: This table reports OLS estimates of equation 4, 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 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, 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, self-reported gender bias and the interaction with students' gender of all these characteristics. We include a control for whether the class has an extended school day and the interaction with the gender of students. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.11: Estimation of the effet of teachers' gender stereotypes on technical technological track

Dependent Variable: Track choice Technical Technological				
	(1)	(2)	(3)	(4)
Fem	-0.242***	-0.257***	-0.403***	-0.397***
	(0.011)	(0.011)	(0.085)	(0.085)
Fem*Bias Teacher		-0.022*	-0.017	-0.017
		(0.012)	(0.011)	(0.011)
Fem*Reported Bias Teacher				0.011
				(0.014)
Std Math grade 6		-0.042***	-0.030***	-0.030***
		(0.007)	(0.007)	(0.007)
Constant	0.310***	0.323***	0.408***	0.407***
	(0.006)	(0.006)	(0.018)	(0.018)
Class FE	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes
Teacher Controls	No	No	Yes	Yes
Obs.	8463	8463	8463	8463
R^2	0.199	0.205	0.220	0.220

Notes: This table reports OLS estimates of equation 3, 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, 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 *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.12: Estimation of the effet of teachers' gender stereotypes

Dependent Variable: Track choice Vocational						
Heterogeneous effects by	**Student Characteristics**				**Interaction time with teacher**	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.017	0.015	0.004	0.017	0.005	0.020
	(0.070)	(0.070)	(0.072)	(0.070)	(0.070)	(0.070)
Fem*Bias Teacher	0.022**	0.022	0.044**	0.019**	0.014	0.035*
	(0.009)	(0.015)	(0.021)	(0.009)	(0.010)	(0.020)
Fem*Bias T*HighEduM		-0.005				
		(0.017)				
Fem*Bias T*Top tercile Math6			-0.048**			
			(0.023)			
Fem*Bias T*Middle tercile Math6			-0.015			
			(0.026)			
Fem*Bias T*Immigrant				0.019		
				(0.027)		
Fem*Bias T*Extended School Day					0.034	
					(0.022)	
Fem*Bias T*Same Math Teacher						-0.016
						(0.023)
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.	8463	8463	8463	8463	8463	8463
R^2	0.210	0.211	0.216	0.211	0.211	0.211

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on the choice of vocational high-school track by observable characteristics of the student and by interaction time with teacher; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 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, 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. Regressions are all fully saturated even if not all interactions are shown in the table. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.13: Estimation of the effet of teachers' gender stereotypes on retention rate and on the probability of doing the standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable	**Retention Rate**				**Doing Test in Grade 8**			
Fem	-0.027***	-0.037***	-0.048***	-0.042	0.014***	0.022***	0.006	0.001
	(0.005)	(0.005)	(0.011)	(0.035)	(0.005)	(0.005)	(0.007)	(0.033)
Fem*Teacher Bias		0.004	0.004	0.002		0.002	0.004	0.005
		(0.005)	(0.005)	(0.005)		(0.005)	(0.005)	(0.005)
Std Math grade 6		-0.053***	-0.042***	-0.042***		0.040***	0.025***	0.024***
		(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
Constant	0.060***	0.070***	0.064***	0.065***	0.939***	0.932***	1.051***	1.052***
	(0.002)	(0.003)	(0.010)	(0.010)	(0.002)	(0.003)	(0.007)	(0.007)
Obs.	9837	9837	9837	9837	9837	9837	9837	9837
R^2	0.099	0.136	0.153	0.154	0.175	0.200	0.390	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 3, where the dependent variable is the retention rate in columns 1-4 and the probability of doing the standardized test score in grade 8 in columns 5-8; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 551. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, 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 *** indicate significance at the 10%, 5% and 1% percent level respectively.

B Appendix: Gender Implicit Association Test

We invite teachers to complete a seven-block IAT that was following the schematic overview presented in Figure B.1. Half of the teachers, randomly selected at individual level, completed the IAT, as presented in Figure B.1, while the other half completed the IAT with the blocks in the following order: 1, 5, 6, 7, 2, 3, 4 ("order incompatible" IAT). In Table A.3, we check the influence of order of blocks on the IAT score. On average, there is a small in the IAT score between individuals that perform the order compatible and incompatible test. Hence, in all regression where there are no class (and therefore teacher) fixed effects we control for the order of IATs. The stimuli presented within each category are summarized in Figure B.2. The improved IAT scoring procedure is used to calculate the bias from the reaction time of individuals to different associations and this method has been shown to be better account for method variance (Greenwald et al., 2003).

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

Figure B.1: Schematic overview of the Gender Implicit Association Test

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

Figure B.2: Category Labels and Stimuli for the Implicit Association Tests

C Appendix: Implicit Bias of Italian Teachers

In this appendix, we report the analyze the role of the implicit gender bias of Italian teachers. Table C.1 summarizes the main observable characteristics of Italian teachers used for this analysis: 91% of teachers are female, almost all have a full time contract and on average almost 24 years of experience as teachers. Female Italian teachers have a stronger gender bias, as measured by IAT score and they are faster in associating own gender with the subject they teach, as seen in Table C.2. Finally, in Table C.3, we report the correlation between student characteristics and teachers' implicit gender bias, measured by IAT score.

The gender gap in reading is reversed compared to the one in math: females are 0.19 standard deviation better in Italian compared to males at the end of the middle school and the gap is increasing of 0.079 standard deviation from grade 6 to 8, as shown in the first two columns of Table C.4. This difference is similar to the one in most OECD countries (Fryer Jr and Levitt, 2010). However, teachers' gender stereotypes are not affecting this gap. In the subsequent columns, we investigate the impact on the gender achievement gap in reading of Italian teachers, including the same controls at student and teacher level as in Table 5. The effect of interest is close to and indistinguishable from zero. The bias of Italian teachers does have impact neither on males nor on females, neither on reading nor on math achievements (as shown before in Table A.8). Indeed, Table C.5 examine the effect comparing students of the same gender enrolled in the same school in the same year (as in equation 4) and shows that there is no impact on both genders when controlling for teachers' characteristics that are crucial when there are only school fixed effects.

Table C.1: Summary Statistics from Italian Teachers' Questionnaire

Family and education					
Female	431	0.91	0.29	0.00	1.00
Born in the North	414	0.74	0.44	0.00	1.00
Age	412	51.37	7.44	33.00	66.00
Children	431	0.73	0.45	0.00	1.00
Number of children	308	1.79	0.82	0.00	5.00
Number of daughters	308	0.81	0.74	0.00	4.00
Low edu Mother	380	0.52	0.50	0.00	1.00
Middle edu Mother	380	0.38	0.49	0.00	1.00
High edu Mother	380	0.11	0.31	0.00	1.00
Advanced STEM	425	0.00	0.00	0.00	0.00
Degree Laude	369	0.29	0.46	0.00	1.00
Job characteristics					
Full time contract	417	0.99	0.11	0.00	1.00
Years of experience	417	23.70	9.59	2.00	43.00
Math Olympiad	425	0.00	0.00	0.00	0.00
Update Courses	425	0.93	0.25	0.00	1.00
Satisfy with teacher job	414	3.89	0.88	1.00	5.00
Implicit bias					
IAT Gender	431	0.39	0.39	-0.78	1.29
IAT Race	427	0.46	0.26	-0.30	1.04
Self-reported explicit bias					
WVS Gender Equality	408	1.07	0.31	0.00	2.00
Gender Dif Innate Ability	402	1.43	0.84	1.00	5.00
Reason GenderGap: Interest for STEM	368	2.65	1.00	1.00	5.00
Reason GenderGap: Predisposition for STEM	342	2.16	1.05	1.00	5.00
Reason GenderGap: Low self-esteem	401	2.54	1.09	1.00	5.00
Reason GenderGap: Family support	400	2.94	1.07	1.00	5.00
Reason GenderGap: Cultural Stereotypes	398	2.17	1.17	1.00	5.00
Reported gender bias	287	0.06	1.02	-1.32	2.20
Boys better in Invalsi	334	0.10	0.30	0.00	1.00
Girls better in Invalsi	334	0.53	0.50	0.00	1.00
Gender Equal in Invalsi	334	0.37	0.48	0.00	1.00
Observations	431				

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.4. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

Table C.2: Correlation between Italian teachers' characteristics and Gender IAT Score

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: raw IAT score of Italian teachers						
Female	0.540***	0.550***			0.578***	0.592***
	(0.071)	(0.069)			(0.072)	(0.106)
Born in the North		-0.061*			-0.056*	-0.017
		(0.033)			(0.033)	(0.052)
Age		-0.045			-0.046	-0.034
		(0.029)			(0.029)	(0.043)
Age sq.		0.000			0.000	0.000
		(0.000)			(0.000)	(0.000)
Children		-0.004			-0.012	-0.038
		(0.046)			(0.041)	(0.055)
Daughters		0.081			0.063	0.073
		(0.059)			(0.055)	(0.072)
Middle edu Mother		-0.009			-0.006	-0.049
		(0.044)			(0.044)	(0.061)
High edu Mother		0.014			0.008	0.011
		(0.048)			(0.050)	(0.061)
Degree Laude			0.006		-0.005	0.015
			(0.036)		(0.036)	(0.046)
Full time contract			0.073		0.170**	0.103
			(0.064)		(0.073)	(0.134)
Satisfy with teacher job			0.017		0.019	0.020
			(0.024)		(0.021)	(0.032)
Explicit Bias				-0.016	0.009	0.014
				(0.022)	(0.020)	(0.027)
Constant	-0.091	1.112	0.085	0.413***	0.668	0.451
	(0.071)	(0.750)	(0.158)	(0.028)	(0.746)	(1.116)
School FE	No	No	No	No	No	Yes
Obs.	431	431	427	431	427	427
R^2	0.188	0.205	0.094	0.030	0.279	0.415

Notes: This table reports OLS estimates of the correlation between Italian 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. 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, "Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. We include the education of the mother of the teacher, whether the degree was achieved with laude, the type of contract and other administrative responsibilities within the school. Check Table **??** for the same correlation for the sample of all Italian teachers available. We include the order of IATs for Italian 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 C.3: Exogeneity of assignment of students to Italian teachers with different stereotypes

Dependent Variable: Italian Teacher implicit gender bias (standardized)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fem	0.018	0.035	-0.018	0.022	0.014	0.024	0.346*
	(0.013)	(0.025)	(0.024)	(0.015)	(0.014)	(0.171)	(0.205)
Fem*HighEduMother		-0.029				-0.038	0.012
		(0.035)				(0.031)	(0.049)
HighEduMother		0.046*				0.052**	0.000
		(0.027)				(0.023)	(0.029)
Medium Occupation Father			-0.002			-0.010	0.062
			(0.028)			(0.026)	(0.041)
Fem*Medium Occupation Father			0.040			0.047	-0.043
			(0.034)			(0.032)	(0.050)
High Occupation Father			0.045			0.038	-0.029
			(0.032)			(0.030)	(0.061)
Fem*High Occupation Father			0.068*			0.079**	0.078
			(0.041)			(0.039)	(0.091)
Fem*Immigrant				-0.018		-0.015	-0.017
				(0.038)		(0.040)	(0.083)
Immigrant				-0.008		-0.016	-0.053
				(0.034)		(0.033)	(0.060)
Fem* Std Math grade 6					0.018	0.014	0.041*
					(0.016)	(0.014)	(0.024)
Std Math grade 6					-0.019	-0.027**	-0.079***
					(0.014)	(0.013)	(0.022)
Fem*Std Ita grade 5							-0.046
							(0.036)
Std Ita grade 5							0.029
							(0.022)
Constant	0.142	0.112	0.138	0.143	0.142	-0.688	-2.135***
	(0.096)	(0.098)	(0.101)	(0.095)	(0.097)	(0.558)	(0.714)
School,year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Control	No	No	No	No	No	Yes	Yes
Obs.	8685	8685	8685	8685	8630	8630	1404
R^2	0.396	0.397	0.397	0.396	0.394	0.500	0.784

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 353 in columns 1-6 and 137 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 6" and "Std Ita grade 5" are the standardized test score in grade 6 in math and grade 5 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 55 students we do not observe the math test score 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 C.4: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - class FE regression

Dependent Variable: reading standardized test score in grade 8						
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	0.192***	0.079***	0.078***	0.088***	0.551**	0.550**
	(0.021)	(0.016)	(0.016)	(0.032)	(0.265)	(0.263)
Fem*Bias Ita Teacher			0.010	0.007	0.008	0.008
			(0.015)	(0.014)	(0.015)	(0.015)
Fem*Ita Teacher Fem					-0.012	-0.017
					(0.051)	(0.051)
Fem*Born North Ita Teacher					-0.013	-0.009
					(0.031)	(0.031)
Fem*Reported Bias Ita						-0.008
						(0.017)
Std Ita grade 6		0.729***	0.729***	0.693***	0.690***	0.690***
		(0.016)	(0.016)	(0.018)	(0.018)	(0.018)
Constant	-0.004	-0.059***	-0.059***	-0.224***	-0.223***	-0.223***
	(0.011)	(0.008)	(0.008)	(0.024)	(0.024)	(0.024)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	Yes
Teacher Controls	No	No	No	No	Yes	Yes
Obs.	8685	8685	8685	8685	8685	8685
R^2	0.184	0.591	0.591	0.603	0.604	0.604

Notes: This table reports OLS estimates of equation 3, where the dependent variable is Italian standardized test score in grade 8; the unit of observation is student i, in class c taught by Italian teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 226. The number of fixed effects (classes) is 289. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between reading standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, degree with honor, update courses, age, type of contract, education of the teacher' mother and self-reported gender bias. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table C.5: Estimation of the effect of teachers' gender stereotypes on reading standardized test score in grade 8 - school FE regression

Dependent Variable: reading standardized test score in grade 8						
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	0.196***	0.090***	0.089***	0.099***	0.595**	0.595**
	(0.024)	(0.018)	(0.018)	(0.032)	(0.250)	(0.248)
Fem*Bias Ita Teacher			0.005	0.001	0.005	0.006
			(0.014)	(0.014)	(0.014)	(0.014)
Bias Ita Teacher			0.039**	0.036*	0.020	0.020
			(0.019)	(0.019)	(0.018)	(0.017)
Female Ita Teacher					0.145*	0.147*
					(0.076)	(0.076)
Fem*Female Ita Teacher					-0.028	-0.035
					(0.055)	(0.054)
North Ita Teacher					-0.017	-0.017
					(0.032)	(0.031)
Fem*North Ita Teacher					-0.002	0.002
					(0.030)	(0.030)
Fem*Reported Bias Ita Teacher						-0.008
						(0.017)
Reported Bias Ita Teacher						0.005
						(0.018)
Std Ita grade 6		0.720***	0.721***	0.678***	0.676***	0.675***
		(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
Constant	0.005	-0.069***	-0.062***	-0.230***	-0.246	-0.253
	(0.021)	(0.017)	(0.024)	(0.032)	(0.169)	(0.169)
School, year FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	Yes	Yes
Teacher Controls	No	No	No	No	Yes	Yes
Obs.	8685	8685	8685	8685	8685	8685
R^2	0.114	0.543	0.544	0.559	0.562	0.562

Notes: This table reports OLS estimates of equation 4, where the dependent variable is reading standardized test score in grade 8; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at Italian teacher level in parentheses; the number of clusters is 226. The number of fixed effects (school by cohort) is 146. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between reading standardized test score in grade 6 and students' gender. Teacher controls include teacher gender, place of birth, degree with honor, update courses, age, type of contract, education of the teacher' mother, self-reported gender bias 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.