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

Michela Carlana†

August 2017

JOB MARKET PAPER

LATEST VERSION HERE AND APPENDIX HERE

Abstract

I explore the influence of teachers' gender stereotypes in affecting student performance, combining administrative data and original first-hand questionnaires on both teachers and students. I 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 negative impact on own assessment of math ability: it 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 failing to achieve their potential.

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

*I am grateful to Alberto Alesina, Eliana La Ferrara 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, Giampaolo Lecce, Andreas Madestam, Valerio Nispi Landi, Paolo Pinotti, Paola Profeta, Jonathan de Quidt, Dan-Olof Rooth, David Stromberg, Jackob Svensson, Guido Tabellini, Marco Tabellini, Anna Tompsett and all seminar participants of Oxford Development Economics Workshop 2017, 32nd AIEL Conference, Human Capital Workshop 2017 Stockholm School of Economics, IIES Brownbag and 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 indebted to Gianna Barbieri and Lucia De Fabrizio (Italian Ministry of Education, Statistics), Patrizia Falzetti and Paola Giangiacomo (Invalsi) for generous support in providing the data. I am grateful to all principals and teachers of schools involved in this research for their collaboration in data collection. I thank Pamela Campa for providing World Value Survey data on Italian provinces. This research project was approved by the Ethics Committee of Bocconi University on 14th September 2016.

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

1 Introduction

Over the last century, the narrowing of gender differences in labor market participation and educational outcomes has been impressive, up to a reversal of the gap in school attainment in many 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 between men and women, which may be more or less accurate but often hold a *kernel-of-truth* (Bordalo et al., 2017). For instance, gap in math performance between males and females, as measured by standardized test scores, is still far from being closed in most societies[1]. However, gender differences in math vary substantially across countries and increase dramatically throughout the educational career of students. 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, I explore whether exposure to stereotypes can causally affect math achievements and track choice of boys and girls.

I focus in particular on the influence of teachers' gender stereotypes in affecting student performance, combining administrative data and original first-hand questionnaire on students and more than 1.400 teachers in Italy. I 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 difference in the additional gap in math performance between boys and girls generated during middle school would be 34 percent smaller if teachers had one standard deviation lower implicit stereotypes. 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, I 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 failing to achieve their potential. Teacher bias fosters low expectations about own math ability and, also through this mechanism, underperformance of individuals vulnerable to the gender stereotype.

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-

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

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 therefore 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 or intention to harm the stigmatized-group (Nosek et al., 2002). In particular, I may expect that teachers do not explicitly endorse gender stereotypes and want to help all their students to achieve what is the best according with their perception. However, students may not be equally encouraged to fulfill their potential if teachers fail to recognize it due to implicit bias. Indeed, research in social psychology shows that elementary school teachers with more gender stereotypes attribute higher math talent to male students than female students with comparable ability (Tiedemann, 2002). Bias can operate through gender specific quality and time of interaction between teachers and students or more subtle forms of “encouragement” to consider own math achievements as the result of talent for boys and effort for girls. 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.

Building on the recent development of economic literature studying discrimination, I 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 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 laboratory experiments (Nosek et al., 2009; Reuben et al., 2014; Glover et al., 2017) 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). 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[2]. 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[3]. 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 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 associations with future performance of students. Compared to other measures of teachers' bias, the Implicit Association Test has two main advantages. First, it does not suffer from social desirability bias that may be an issue in self-reported measures of stereotypes. Second, student performance is not used to create the measure of teachers' bias, solving potential concerns in the construction of the stereotype measure related to gender-class specific factors, which may have a stronger impact on grades given by teachers compared to blinded test-scores.

Girls outperform boys in reading and in other dimensions of school attainments. So why should I 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 gap in math skills between males and females is of first-order importance to explain the enduring gender gap in

[2]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.

[3]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.

performance and the underrepresentation of women in leadership position and among science, technology, engineering, and math (STEM) workforce.

This paper makes three substantial contributions. First, it exploits a newly-available administrative dataset on students merged with first-hand questionnaire information on their teachers to analyze the role of educator stereotypes in affecting student performance in math and track choice. It offers new micro-level evidence that exposure to cultural bias has a relevant role in affecting gender gap in math performance, in line with Guiso et al. (2008). Second, this dataset allows to investigate precisely the impact of *implicit* bias, contributing to the recent literature in economics that has underlined its potential relevance in real-world settings (Glover et al., 2017; Bertrand and Duflo, 2016). Furthermore, the richness of the data offers evidence of the correlation between implicit bias, as measured by the Implicit Association Test, and background characteristics of individuals, as cultural stereotypes in the place of birth of teachers and field choice. Lastly, it provides evidence that teachers' bias influence self-assessment of math ability and through this channel it influences the underperformance of girls in math. In this perspective, the conceptual framework described in next section offers an intuitive explanation of this mechanism.

After a stylized model of the influence of teachers' stereotypes on students' math performance presented in Section 2, the paper proceeds as follows. Section 3 explains the setting analyzed, providing information on the Italian cultural and institutional background. 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 entrenched in most societies and female students are potentially vulnerable of being judged by the predicament that "girls are not good at math". We may expect that teachers in OECD countries do not explicitly endorse gender stereotypes and do not want to intentionally harm any student. However, gender stereotypical beliefs, rooted in own experience since childhood, may affect learning of pupils (at least) through two channels: self-perception of own math ability of children and interaction between teachers and students.

Indeed, if the student perceives higher bias toward her own group, she may decrease assessment of her 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 by beliefs 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 teacher have an impact on gender gaps in math performance. This conceptual framework is related to the model of Coate and Loury (1993), in which bias impede skills development. Furthermore, it is consistent with the idea of stereotype threat (Steele and Aronson, 1995) developed in 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 steps taken in this direction has been Fryer et al. (2008) that find 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 spend less time (in terms of either quantity or quality) interacting with girls, especially those poorly performing. Teachers may choose to allocate more time or tailor math classes to the learning of boys and top-performing girls since they are more likely to attend a STEM track during high-school. Finally, teachers with higher bias may simply fail to recognize their talent in math related fields. Therefore, they may set a lower bar for their learning and not encourage them to fulfill their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983).

I develop a simple conceptual framework to analyze the impact of teachers' gender stereotypes on effort of students, as mediated by student perception of own ability and teacher behavior toward the pupil (encouragement and/or time investment). In this simple 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, I observe improvements in achievement test scores (P) and not directly effort (e)[4], but I assume for simplicity that the derivative of performance with respect to effort is positive ($P_e > 0$) and in this section I focus on the choice of the latter.

[4]Even ideally having information about the number of hours studied, it is not clear that this is necessarily a better measure of effort since the quality of time use is also essential in the learning process

I extend this framework in two directions to capture the influence of teacher stereotypes toward her own gender (s_t) on the behavior toward the pupil (β_{t_i}) and students' belief formation about her own ability level (α_i).

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

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

where u is differentiable and the sufficient conditions for a local interior maximum hold. In this model, I do not introduce parametric assumptions on the utility function.

Students beliefs about own ability in math is a function of own true ability a_i and teachers' gender stereotypes, defined as follows: $\alpha_i = f^i(a_i, s_t)$. The impact of the bias on own self-perception is an individual specific function: students with higher vulnerability to the stereotype that "girls are not good at math" will be more negatively impacted by teacher stereotype. This simple framework is flexible enough to capture heterogeneous perception of teachers' bias among classmates. For instance, boys may not perceive gender stereotype in math since they do not belong to the stigmatized group. I assume that students' beliefs about own ability is a decreasing function of teachers' stereotype, i.e. $\alpha_s \leq 0$.

Teacher behavior toward the pupil is captured by: $\beta_{t_i} = h^i(s_t)$, where I assume that teachers with higher stereotypes are less supportive toward member of the stigmatized group, such that $\beta_s \leq 0$.

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

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

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

Effort and perceived ability are often considered as complementary in the education production function (i.e. $k_{e\alpha} > 0$), so that a higher self-assessment of own capacities enhances the motivation to exert effort (Bénabou and Tirole, 2002). Hence, if also effort of the student and supportive attitude of teachers are complementary, then higher educator stereotypes will decrease the level of effort in equilibrium ($e_s^* \leq 0$). However, if the student increases diligence as a reaction to a negative stance of the math teacher, then the impact of stereotypes on effort is not clearly defined and it depends on sign of the numerator in equation (2). As suggested also by Dee (2014), $e_s^* \geq 0$ is likely for instance if individuals of the stigmatized group consider the stereotype strongly improper and react with an "I'll show you are wrong" attitude. In the context of gender stereotypes, it would imply that talented female students may increase the level of effort when they interact with teachers with stronger bias in order to disprove the negative belief.

In the the empirical analysis, I analyze how improvements in achievement test scores are affected by teacher stereotypes. Assume two students, with the same gender, family background and math performance, are quasi-randomly assigned to two different teachers, respectively with stereotypes s_{t_i} and s_{t_j}, such that $s_{t_i} < s_{t_j}$. Then, if effort is complementary of both own perceived ability and teacher attitude, according with this conceptual framework, the optimal level of effort (and therefore performance) of the student decreases with teachers' stereotypes. Unfortunately, I do not observe data on gender specific investment or interaction in the classroom between professors and students. However, I do show the importance of the channel related to self-stereotypes and how they are influenced by educators' own gender bias.

3 Setting

3.1 Culture and Female Labor Force Participation

Italy is a country with low labor market participation of women: female employment rate aged 15-64 years in 2016 was 47 percent, with substantial geographic variations within the country. Only 31 percent of women in the South of Italy were employed, while in the North around 58 percent were working, similarly to the average of OECD[5]. The geographical difference in employment rate of men in Italy is smaller compared to that of women: male employment rate is 74 percent in the North and 55 percent in the South, suggesting that the difference observed in female employment rate is not entirely driven by labor market conditions and factors unrelated to gender.

I study the influence of teachers' gender stereotype in a sample of Italian schools located in the North of Italy (Milan, Brescia, Padua, Genoa and Turin), where around 40 percent of math teachers are born in the South of the country. Gender stereotypical beliefs are rooted in cultural traits, transmitted fairly unchanged from generation to generation (Guiso et al., 2006). An increasing literature has shown the impact of culture and attitudes on labor force participation of women, using different approaches to deal with the endogeneity of cultural traits (Giavazzi et al., 2013; Fernández, 2007). The epidemiological approach introduced by the latter paper has been exploited by Nollenberger et al. (2016) to show that among second generation immigrants,

[5]The average employment rate of females in 2016 in OECD countries is 59.5 percent (Source: OECD and Istat).

the higher the degree of gender equality in the country of origin, the higher the performance in school of girls with respect to boys. Within Italian regions, beliefs about the role of women in the family and workplace are heterogeneous and seems to be a determinant of female labor force participation, as shown by Campa et al. (2010) using data on attitudes from the World Value Survey. The culture in the place of birth of educators may have influenced their perception of the role of women in the workplace and therefore their gender stereotypes.

3.2 Institutions: the Schooling System

In the Italian educational system, middle school lasts three years from grade 6 to 8, after which students self-select into three different tracks: academic oriented high-school ("liceo"), technical high-school and vocational high-school. Each type of school is divided in several subtracks: the academic oriented track can be specialized in either scientific, humanistic, languages, human sciences, artistic or musical subjects, the technical track can be focused on technological or economic subjects, while the vocational track can have different core subjects as for instance hospitality training, cosmetics and mechanical workshop. Students are free to choose the high-school they like the most, with no restriction on the track 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.

Students at the beginning of middle school are assigned to classes and they stay with the same peers for three years[6]. A crucial aspect for this paper is the assignment of students to teachers, within the same school. The general class formation criteria are established by an Italian law and details are specified by each school council[7]. The information on class formation criteria for each school is available on a formal document on the website of the institution[8]. The general criteria mentioned directly by most schools are equal allocation of students across classes according with gender, nationality, disability, socio-economic status and ability level (as reported by the elementary school). I also collect new information directly from the principal on how classes are formed, which are described in details in Appendix C. Headmasters report that the most important aspect in the class formation process is the homogeneity across and heterogeneity within group of students of the same school. 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, while all schools in my sample are from the North. What is important for my analysis is that I can also test whether this intention of the principals is

[6]There are only few exceptions: students may be transfer to a different school or being retained (less than 5% of students).

[7]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.

[8]This document is called "Plan of Education Offer"("Piano dell'Offerta Formativa")

confirmed by the allocation of students to classes in my sample (see section 5.1.2).

Teachers are assigned to schools by the Italian Ministry of Education and they are all paid equally in a centralized system. Teachers' allocation across school is determined by seniority: teachers with more years of experience tend to move away from disadvantaged backgrounds (Barbieri et al., 2011). Each class is assigned by the principal to a math and Italian teacher among those available in the school and they usually follow students from 6 to 8 grade. Every week, students spend at least 6 hours with the math teacher and 5 hours with the Italian teacher[9].

Standardized test score in math and reading are administered in grade 2, 5, 6[10], 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 of the final score of students at the end of middle school. However, this final grade has no relevant impact for the enrollment in high-school or for the future 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 they 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 by other factors as diligence, effort and improvements over time. Grades are given in a scale up to 10, where the pass grade is 6.

4 Data and Descriptive Statistics

During September 2016, I invited 156 middle schools to take part in a research project regarding "The role of teachers in high-school track choice", out of which 91 accepted and provided all information necessary for my study. The sample was designed including all schools of the provinces of Milan, Brescia, Padua, Genoa and Turin with more than 20 immigrants in the scholastic year 2011-12 enrolled in grade 6[11].

I use four sources of data: teacher survey data, student survey data, administrative infor-

[9]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

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

[11]More precisely, in 103 schools we obtain the authorization of the principal to administer the survey to teachers, but only 91 principals completed (without mistakes) the formal authorization to give me access to data from the National Center for the Evaluation of the Italian Educational System (INVALSI). In 2 cases, the principal explicitly stated they did not want to give access to INVALSI data. In most of the cases, the authorization (with all correct data) was not sent in time for the extraction of data from INVALSI. Finally, the number of schools according with 2011 data were 145. However, some of them where divided in different institutions and we follow all of them.

mation from the Italian Ministry of Education (MIUR) and from the National Center for the Evaluation of the Italian Educational System (INVALSI). I collected directly detailed information on teachers, including *implicit bias* measured by the Gender Implicit Association Test (IAT), and on students' self-assessment of own ability in different subjects. Administrative data from MIUR 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.

4.1 Teacher Survey

From October 2016 to March 2017, I 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[12]. The survey is divided into two parts: the Implicit Association Test (IAT) and a questionnaire.

4.1.1 Gender Bias and IAT Test

In this research, the main focus is on *implicit* gender bias. Indeed, 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 (Bertrand et al., 2005). Examples of gender stereotypical beliefs that operate even without awareness or intention to harm the stigmatized-group 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. Similarly, math teachers may unintentionally set a lower bar for females' achievements, by being more generous toward them in grading, as emerges in most countries from the comparison of blind and no-blind test scores. Social psychologists have underlined that teachers with more conservative gender norms are more likely to believe that the performance of male children is attributed to talent compared to female children with similar math ability, with consequences on the language and interaction

[12]Only 4 math teachers, started the questionnaire and then did not finish it since they claimed either that they were not expecting such a long survey or that they could not understand the scope of the Implicit Association Test.

teacher-student (Tiedemann, 2002).

Fully acknowledging the importance of considering implicit bias in economic literature studying discrimination, I 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[13]. The IAT requires the categorization of words to the left or to the right of a computer or tablet screen and it provides a measurement of the strength of the association between two concepts (as for instance gender and scientific/humanistic subjects). Enumerators 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 [14]: 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)[15]. IAT

[13]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.

[14]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.

[15]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).

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, 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 I directly test the impact of implicit stereotypes on the stigmatized group.

The measure of bias I collect is strongly related to the context of my study: school performance and choice. Coherently with the purpose of our paper, I 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[16]. Furthermore, individuals complete the survey in the presence of a enumerator and therefore we are sure of the identity of the teacher who completed the survey[17].

4.1.2 Teachers' Questionnaire

After the Implicit Association Tests, numerators invite teachers to complete a questionnaire with detailed information about family background of teachers (age, parents' education, place of birth, age and sex of children, ...) and career related aspects (type of contract, years of

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

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

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, ...). Furthermore, they are also asked questions about explicit forms of bias, as for instance beliefs about gender differences in innate math ability and the standard Word Value Survey question: *"When jobs are scarce, men should have more right to a job than women"*. We 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) and an assessment of the average performance of their students in the standardized test score, by gender, including also a question in which they asked to reveal how sure they feel about their answer. Finally, we asked teachers which factors they 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). Detailed information about the questionnaire administered to teachers are available in the Appendix B. Enumerators collected the allocation of teachers to classes from the scholastic year 2011-12 to the scholastic year 2016-17, in order to merge teacher and student data. I double checked all these information using data provided directly by schools and their websites.

I 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 market. Factor analysis is performed using polychoric correlation matrix since variables are ordinal and separately for math and Italian teachers. Factor loadings are presented in Table A.1. 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.

4.2 Administrative Data

Thanks to the authorization of each schools' principals, I was allowed to obtain information from the Italian Ministry of Education and from the National Center for the Evaluation of the Italian Educational System (INVALSI). I collected individual level data for three cohort of students enrolled in grade 6 between scholastic year 2010-11 and 2012-13 [18]. The data available include math and reading standardized test score in grade 6 and 8 from INVALSI, together with information from their parents' education and occupation. Data from the Italian Ministry of Education include baseline individual information (date and place of birth, gender, citizenship)

[18] I have information on later cohorts as well, but it is not complete since the test score in grade 6 was not administered after the scholastic year 2012-13.

and 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.

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[19]. Students in grade 8 in 2014, around two months before the end of middle schools, are asked to complete a survey about their track choice. In particular, they need to mention all subjects they will learn during high-school and to report their self-belief about own ability in each subject. The potential choices to that answer were: "good", "mediocre", "scarce". In all high-school, mathematics is taught and therefore most students report their self-assessment of math ability.

4.4 Descriptive Statistics

4.4.1 Math Teachers

The dataset includes 537 math teachers, 855 Italian teachers and 31 teachers of other subjects. The main focus of this paper is on the impact of math teachers gender stereotypes on the performance in the subject they teach. Among these 537 teachers, we include in the main analysis 301 teachers that were hired by the same school in the scholastic years between 2010-11 to 2014-15 and for which we obtained the authorization from the principal to access data on standardized test scores of students[20]. Table A.3 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, 40 percent less likely to have full-time contract, have 12 years less of experience in teaching and they are 17 percent less likely to have children. However, as it can be clearly seen also from Figure **??**, not only the average, but also the entire distribution of implicit gender bias of the matched and not-matched teachers is extremely close (exact p-value of Kolmogorov-Smirnov: 0.946).

Table 1 reports descriptive statistics of math teachers' information. Most of teachers are females (84%), they are on average 52 years old with 23 years of experience in teaching and

[19]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 I have complete information on standardized test score and teachers' bias. The specific question exploit in this paper was collected only for the control group and therefore not used in the paper by Carlana et al. (2017).

[20]As specified in section 4, for 12 schools we did not obtain this authorization on time or there was a mistake in the authorization form. We lost some observations because some schools changed the official code (called "meccanografico") over the years of our sample and INVALSI guarantees access to data only for school codes whose principal has signed the authorization.

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 related subjects. In the last part of Table 1, I report the summary statistics of explicit bias questions described in details in Appendix B. The variation in the answers on the equality of access to labor market of men and women and about innate gender difference is ability is low, potentially also due to social desirability bias: for instance, less than 2% of the interviewed teachers affirm to agree with the statement that women have less right to jobs than men when opportunities are low. It may be difficult to obtain revealed bias, given the widespread explicit rejection of stereotypes and a related reluctance of participants in revealing their bias, especially if interviewed as "teachers" in the presence of enumerators.

Math teachers are slightly gender biased: indeed, a positive IAT score indicates a stronger association between males with scientific subjects and female with humanistic subjects. For ease of interpretation of our results, I standardize the IAT score to have mean zero and variance one throughout the paper. Considering the thresholds typically used in the social psychological literature[21], 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[22] in a similar Gender-Science test have an average score of 0.40 (SD 0.40): the score of math teachers is on average lower than this sample (mean 0.09, SD 0.37, as shown in Table 1), while Italian teachers are very close to it (mean 0.39, SD 0.39, as shown in Table E.1). Interestingly, the great majority of math teachers are females and this may have important implication for the association of scientific subjects with gender.

4.4.2 Correlation between implicit bias and individual characteristics

The richness of the data collected allows to associate individual level characteristics of teachers with the results from the Implicit Association Test (IAT) in order to dig deeper into the determinants captured by reaction time to stimuli. Table 2 shows the correlation between math teacher IAT score and their characteristics. Women teaching math are significantly less biased in associating own gender with STEM and this aspect explain a substantial portion of the low average IAT score for math compared to Italian teacher. Figure 2 plots the entire distributions

[21] 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.

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

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 columns 3-6 (Panel A), I show the association with age, education of teachers' mother and whether teachers have children or specifically daughters. Among this group of adults, implicit stereotypes do not seems to change in the different age groups. Teachers with mothers that graduated from high-school seems to be slightly less biased, even if the effect is imprecisely estimated. Finally, having children, and in particular daughters, do not significantly impact on gender stereotypes. In Panel B, columns 1 and 2, I correlate the IAT score with qualifications of the teacher (type of degree and whether the degree was achieved with honor), finding negative point estimates despite high standard errors. Another rough proxy of potential quality of teachers is related to having the tenure (which is associated with higher experience in teaching), attending update courses and being the professor in charge of math Olympiads in the school[23]. Also in these cases, point estimates are small and indistinguishable from zero.

Exposure to cultural norms is associated with the IAT score. In column 2 of Table 2 (Panel A), I correlate the implicit bias with the place of birth of teachers. Around 35 percent of math teachers in this sample are born in the South where gender norms are stronger, as shown for instance by Campa et al. (2010) using World Value Survey data at Italian provincial level. In Panel C, I further investigate how implicit association are correlated with individual level beliefs and cultural norms in the place of birth. Participation of women in the labor force in the province of origin of teachers is correlated with the main stereotype measure exploit in this paper[24]. In column 2 and 3, Panel C, I adopt the standard definition of culture based on individual preferences measured by World Value Survey data at Italian provincial level. I find that the answer to the question on the relative rights of men and women to paid jobs when the latter are scarce is significantly correlated with IAT score when considering the average in the province in which the teacher was born. The correlation is low and indistinguishable from zero when considering the answers given by teachers during the survey I administered. We may suspect that either there is a social desirability bias in self-reported measured when professors are interviewed in the school or they believe they are not affected by the bias they were exposed to during childhood. In general, the correlation between explicit self-reported measures and IAT score is low. In column 5, Panel C, I correlate implicit and reported bias and I find a weak positive correlation (not statistically significant). This result is not surprising in light of social psychology literature,

[23]In each school, usually only one professor is in charge of math Olympiad and anecdotally she is highly motivated and passionate teacher.

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

where implicit often differ from explicit and self-reported stereotypes (Lane et al., 2007; Nosek et al., 2002).

Appendix Table A.4 shows jointly all correlation presented in separate regressions in Table 2 Panel A and B, for the sample of teachers matched with students before 2016 and for all math teachers interviewed. Interestingly, the results are substantially invariant: gender of the teacher and place of birth are the two most relevant aspects in affecting IAT scores in all specifications. Using the whole sample of math teachers, the degree obtained has a statistically significant impact on stereotypes: teachers with an advanced STEM degree have on average lower implicit bias.

4.4.3 Students

Table 3 reports balance tables of students' information obtain from administrative data and from the questionnaire with original first-hand data. I restrict the sample to students with information available on the standardized test score in grade 6 and 8 and for which I 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. Appendix D describes in details the sample selection and potential attrition issues.

In our sample, 50% of students are males and boys and girls are balanced in terms of baseline characteristics related to place of birth, generation of immigration, parents' education and occupation. Test scores are standardized to have mean zero and standard deviation one per subject and year in which the test was taken. Females at the beginning of middle school are lagging behind of 0.19 standard deviations in math and ahead of 0.13 standard deviations in reading, with respect to males.

In the same table, I also report the raw gender differences in outcomes. The high-school track choice in this sample is comparable to the average national choices in those years: females are almost 10 percentage points less likely 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[25]. Finally, from the original first hand information available for a sample of students, I observe that on average there are no gender differences in assessment of ability, but females are 9 percentage points less likely to consider themselves good at math and boys are 5 percentage points less likely to consider

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

themselves good in Italian.

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. I exploit two identification strategies. The former is aimed at investigating the impact of teacher bias on the gender gap within a class, estimating the following equation:

$$y_{ict} = \alpha_0 + \alpha_1(Female_i \times bias_{ct}) + \alpha_2 Female_i + \eta_c + \mathbf{X}_i\rho_1 + \varepsilon_{ict} \tag{3}$$

where y_{ict} is an outcome of student i in class c taught by teacher t in grade 8 (as for instance math standardized test score). $Female_i$ is a dummy variable which assumes value 1 if the student i is a girl and $bias_{ct}$ is the standardized value of the gender implicit bias of the professor teaching in class c in grade 8[26]. I include fixed effects at class level η_c, which absorb the average effect of the bias in class c, and controls for student characteristics $\mathbf{X}_i$ (standardized test score in grade 6, parents education and occupation, immigration status and generation of immigration). Additionally, for robustness, the regression controls for the gender of students interacted with student characteristics $\mathbf{X}_i$ and teacher characteristics $\mathbf{Z}_{tc}$ (as gender, place of birth, type of contract, type of degree, self-reported gender bias, ...)[27]. Standard errors are robust and clustered at teacher level.

Crucially, in this identification strategy all class, teacher and school level characteristics are absorbed by class fixed effects. We can only identify the impact of teacher bias on the class gender gap in the dependent variable, i.e. the interaction between the gender of students and implicit stereotypes of teachers. However, the inclusion of the interaction between gender of students and teacher characteristics ($\mathbf{Z}_{tc}$) is potentially important to partial out potential differential impact on males and females of sex, background and experiences of teachers. Furthermore, they allows to establish whether the impact of teacher stereotypes on gender gap among classmates can be explained (or attenuated) by teachers' observables given the correlation between some teacher characteristics and IAT scores pointed out in Table 2.

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. I expect the estimate of α_1 to be attenuated for the measurement error in the gender IAT score. Indeed, occasion-specific noise may introduce error in the estimation of implicit bias and, as suggested

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

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

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.

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

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

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

All institution level characteristics are absorbed by school cohort fixed effects. The advantage with respect to specification (3) is that we can analyze the impact of teacher stereotypes separately on male students (β_3) and on female students ($\beta_1 + \beta_3$). The drawbacks is that we cannot control for unobservable characteristics at teacher or class level: this specification exploits variation in the level of teacher bias to which students of the same gender in the same school are exposed.

5.1 Identification Issues

5.1.1 Reverse Causality

The measure of teacher gender stereotypes was collected between October 2016 and March 2017 and the time-line of data available also for students is presented in Figure 1. Teacher biased is collected after students in the sample graduated from middle school. The main advantage of this timing choice is that 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 cohorts of boys and girls under analysis. The main potential issue of this approach is reverse causality if teacher bias, as measured by the implicit association test, is affected directly by exposure to students. This would be the case for instance if teachers with higher bias were assigned to groups of females with low potential or males with impressive math abilities.

The IAT is expected to be the combination of two aspects: the former is a trait stable across time capturing the influence of cultural norms and experience throughout own life, while the latter is noise, occasion-specific variation that may be affected by conditions while taking the test and stimuli received by the subject in the period right before the test. The test-retest reliability of IAT is generally considered as satisfactory by social psychology, with a correlation of

0.56 that does not change with the length of time between testing (despite being usually of less than one month in most studies) (Nosek et al., 2007).

Teachers' experience in school may shape their implicit bias. 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. Crucially for our analysis, we do not include in the sample the cohort of student graduating right before the scholastic year in which the test was administered. Each teacher has been exposed on average to 4 classes (around 100 students) after those included in our analysis[28]. Hence, the noise component is unlikely to be affected by the cohort of students analyzed in this paper.

Furthermore, I can provide evidence of a crucial robustness check: if it was the case that recent graduates influence the measure of teacher stereotypes, we would expect a stronger impact on implicit bias on the most recent cohort of students. I will provide evidence in Table A.12 that results are stable for the three cohort I analyze.

Last but not lest, teachers with higher bias did not receive a different "treatment" in the type of students they were assigned to. This aspect is crucial not only for the reverse causality issue in the measurement of teacher stereotypes and it will be discussed in section 5.1.2.

5.1.2 Exogeneity Assumption

Before turning to analyze the empirical strategy, I 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, as emerge from the official documents in the school websites of most schools ("*POF-Piano Offerta Formativa*") and from self-reported information from principals discussed in Appendix C. However, I check whether students baseline characteristics (gender, education and occupation of parents, immigration status, generation of immigration and initial test scores) and the class assignment are statistically independent with a series of Pearson Chi-Square tests, considering both gender together and divided. I find that less than 10% of the tests performed, the p-value is lower or equal than 5%.

However, not all students enrolled in the schools of our sample in the period 2011-2015 are included in the analysis. Indeed, for this paper, we need information of math teacher stereotypes and standardized test scores in grade 6 and 8. The characteristics of the sample are described

[28] 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 around 100 students after the last cohort of students I analyze and before taking the IAT.

in the Appendix D. In Table 4, I provide evidence that the characteristics of students are not systematically correlated with the implicit bias of teachers. I would not be able to obtain causal estimates if teachers with higher gender bias are systematically more/less likely to be assigned to females or to females with specific characteristics in terms of parents' education and occupation, place of birth and ability. I might expect that if parents had control over assignment of their children to teachers, daughters of highly educated mothers would have been less likely to be assigned to more biased teachers, within school. Instead, I see that the difference is not statistically significant and the point estimate goes in the opposite direction. In columns 3, 4, 5 and 6, I analyze the correlation respectively with father occupation, immigration background and for the proxy of ability using standardized test scores in reading in grade 6 and I do not find statistically significant correlation and also very small point estimates. Finally, in the last column, I also include the standardized test score in math in grade 5, before entering middle school, despite the sample size is substantially reduced for data availability issues[29]. The assumption of quasi-random assignment of students in the sample to teachers with different level of gender bias, as measured by the Implicit Association Test, within a school, seems to be supported in the context under analysis. The result is identical when observations are collapsed at teacher level, as shown in the Appendix Table A.5.

5.2 Performance in math

Girls are lagging behind in math compared to their male classmates of around 0.22 standard deviations by the age of 14, a result comparable to several other countries (Fryer Jr and Levitt, 2010; Bharadwaj et al., 2016). Most of the variation in math performance is not coming across classes, but within, coherently with the target in class formation of heterogeneity within groups and homogeneity across groups: the average gender gap without controlling for class fixed effects if anything is slightly smaller (0.21 standard deviations) (see Table 3). As children complete more years of education, the differences between boys and girls gets bigger. The additional gender gap in math generated during the last two years of middle school is around 0.08 standard deviations, as shown in column 2 of Table 5.

Before moving to the causal estimates, Figure A.2 plots the relationship between bias of teachers and math performance of male and female students. Each circle plots the average improvement in math test scores of students assigned to a math teacher with the indicated level of bias, aggregated into bins. The size of the circle indicates the number of observations per bin. These graphs do not remove any fixed effect at individual, class or school level, but it is a simple plot of raw data. Nonetheless these figures tells a similar story compare to our

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

regression analysis: female students are lagging behind when assigned to math teachers with higher implicit bias.

Table 5 shows the effect of teacher bias on gender gap in math performance within the class, presenting the results of estimating equation (3). Classes that are assigned to teachers with one standard deviation higher bias have 0.027 standard deviations 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. The forth column of the table includes students characteristics $\mathbf{X}_i$ and their interaction with gender of the children. Adding these controls does not change the coefficient of interest. In the whole 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.

Despite the level of teacher bias and all characteristics are absorbed by the class fixed effect, as clarified describing equation (3), column 5 includes the interaction between student gender and teacher characteristics $\mathbf{Z}_{tc}$. If anything, the coefficient of interest *Fem*Bias Teacher*, which corresponds to α_1 in equation 3, slightly increases in magnitude when all these interaction effects are absorbed. Observable characteristics of teachers, interacted with students' gender, are not driving the relation between gender gap and teacher bias. I report the coefficient only for the main characteristics of teachers interacted with students' gender, but the effects are mainly small and insignificant for all variables, including age, parents' education, whether she has children or daughters,whether she achieved the degree with *laude*, the type of teaching contract, update courses and appointment as teacher in charge of math Olympics, which potentially capture the commitment and passion to teaching of the professor. 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 compared to their classmates. The impact of teacher gender is coherent with the result of Bharadwaj et al. (2016), but other papers find that having a teacher of own gender helps improving performance, especially at college level (Dee, 2005; Carrell et al., 2010). Finally, having a teacher born in the North of the country does not have an heterogeneous effect on boys and girls. The results are robust to potential confounding aspects considering all information available on professors from their family background to their professional career.

Finally, in column 6, I consider the impact of self-reported gender bias as well. Despite I acknowledge several potential issues in the self-reported measure, as social desirability bias and reverse causality, it is worth noting 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 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 (as shown in Table A.1).

Are biased teachers worst instructors or are they helping boys to learn math? I 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. Furthermore, controls for the amount of math hours per week are included in this specification and interacted with the student gender. Indeed, in almost all schools some classes have an extended school day and they spend more time with all teachers, including the math one. Adding all these controls does not significantly impact on the main results. Explicit gender bias has a negative 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.

5.2.1 Heterogeneous effects

We now examine who are the students mainly affected by teacher bias. 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. Indeed, Table 7 shows that the effect of implicit stereotypes is stronger for the most disadvantaged groups of female students, in term of background characteristics. All columns in this table include class fixed effects, student and teacher controls interacted with the gender of pupils and for comparison column 1 reports the impact of teacher stereotypes as in column 5 of Table 5.

Column 2 investigates heterogeneous 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 0.049 standard deviations among students with low educated mothers and of 0.027 standard deviations among students with higher level of education (at least a diploma), despite the difference is indistinguishable from zero. In the following column, I analyze the impact of teacher bias in the three terciles of the distribution of the standardized test score in grade 6. The effect is stronger for students in the lowest tercile (-0.070, with standard error 0.027) and turning positive, but not statistically significant, only for students in the top of the initial ability distribution in grade 6. Finally, the effect if anything is slightly stronger among immigrants, even if the difference with natives is not statistically significant at usual levels.

Why do girls from more disadvantaged backgrounds suffer the most from the interaction with biased teachers? These girls may be more vulnerable to the stereotype and have a higher risk of confirming negative expectations on their group. One complementary explanation, coherent with the interaction theory (McConnell and Leibold, 2001), is that female students with highly educated mothers or with higher initial level of math achievement may need less interaction with their math teacher in order to avoid lagging behind with their peers. They are more likely to have additional support to believe in their own abilities and alternative role models with respect to their math teachers.

In order to investigate further the second potential explanation, I analyze the heterogeneous effect according with the "quantity" of interaction time between teacher and students, in terms of years of exposure and hours per week. The last two columns of Table 7 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, I exploit the fact that around 20% did not have the same teacher for all three years of middle school. However, for both variables, I do not see a statistically significant pattern. Most likely the impact of teacher gender stereotypes kicks at lower intensive margins and we do not have proxies of the "quality" of teacher- student interaction that would be necessary to further investigate this aspect.

5.3 Self-Stereotypes

One potential mechanism that can explain the impact of teacher bias on math performance of female students is self-stereotyping. Indeed, expectations about their group's suitability for math affect 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 in 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, I clarify the role of this particular mechanism. To test whether teacher activate self-stereotypes, I asked to a sample to students in grade 8 the extent to which they feel comfortable in their ability in different subjects. Table 8 assesses the extent to which bias of teachers affect own assessment of ability, for a sample of around 800 students for which I collected these information (as described in section 4.3). I present results for self-stereotypes in math in Panel A, in reading in Panel B and on average on all other subjects in Panel C.

As shown in column 1, females are 9.4 percentage points less likely to consider themselves good at math (which corresponds to 11% percent lower probability than males), 5.2 percentage points more likely to consider themselves good in Italian (which corresponds to 6% percent higher probability than males), but on average both equally assess their own ability. In classes assigned to math teachers with higher bias, the gender gap in self-assessment of own ability in math 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, controlling for the test score in grade 6 as in our main specification in equation (3). Adding student and teacher level controls interacted with pupil gender do not substantially affect the point estimate of interest as clearly emerges from columns 3 and 4.

Section 5.2 provides evidence that the gender gap in math achievement increases in classes assigned to more biased teacher. Hence, in the last three columns of Table 8, I 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. I examine whether gender gap in own assessment is merely due to higher gender gap in performance at the end of middle school. I find that gap in own assessment is reduced only by less than one third: teacher stereotypes have an additional impact on own assessment of math capabilities, on top of measured ability, that may have detrimental effects for investment choices in STEM education. In the Appendix Table A.7, I show the result of the specification described in equation (4). Results are coherent with the negative impact of teacher bias on self-stereotypes of female students and no impact on male students. All results are robust to the inclusion of controls at pupil and teacher level and their interaction with student gender.

In Panel B and C of both Table 5.2 and A.7, I focus on the impact of math teacher bias on self-assessment in other subjects different from math. More precisely, in Panel B, I focus on Italian, the other main subject taught during middle school in terms of number of hours, and in Panel C, I focus on the average on all other subjects. Female 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) and are coherent in both specification the former including class and the latter school fixed effects. Finally, in the last three columns of Panel B, I control for the standardized test score in Italian in grade 8: as expected, it does not affect the estimate since math teacher stereotypes do not impact gender gap in reading performance. A deeper analysis of the impact on reading test scores or of the gender gap of the Italian teacher is introduced in Appendix E).

5.4 Further Discussion on Mechanisms

In this section, we refer to the conceptual framework presented above and discuss the predictions in the light of the results on the impact of teacher bias on student math performance. The focus is on two main potential mechanisms: self-stereotypes and interaction theory.

Teachers may held erroneous expectations that lead to a self-fulfilling prophecy or they may fail to recognize students' talent and therefore not encourage them to fulfill their potential (Rosenthal and Jacobson, 1968; Cooper and Good, 1983). In both these cases, exposure to bias activates *self-stereotypes* on female students and, through this channel, leads to under-achievement in math compared to own potential. The result on own assessment of math ability are coherent with triggered self-stereotypes activated both consciously and unconsciously in the students. Relatedly, the empirical evidence presented is coherent with stereotype threat model (Steele and Aronson, 1995): individuals with higher risk of conforming to the predicament that "women are bad at math" are those more deeply affected. Indeed, male students are not influenced by the stereotype of teachers and among females those strongly involved are from disadvantaged backgrounds, especially in terms of initial math achievements.

According with the *interaction theory*, individuals with higher implicit bias are more hesitant and they spend less time interacting with members of the out-group (McConnell and Leibold, 2001). 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. In our context, students in the same class are exposed to teachers for the same amount time during taught lessons. The last two columns of Table 7 try to exploit variation on the time of exposure to teachers in terms of hours per week and years with the same teachers, without finding statistically significant differences at these margins. Unfortunately, we do not have measures of the "quality of interaction" between teachers and student by gender. We cannot exclude that teacher with more bias try spend more time at improving male achievements and supporting their learning to the detriment of females. However, even if this is the case, boys do not benefit from being assigned to a biased teacher neither in terms of higher performance nor in terms of more self-confidence.

Finally, a third theory could be consistent with the negative impact of teacher bias on female student math performance. According with the (*animus theory*), teachers may dislike female students, treat them badly or 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. Furthermore, in the appendix G we provide evidence that teacher favor female students in math grading, comparing blinded and no-blinded scores, as emerges in several other countries (Lavy and Sand, 2015; Terrier, 2015).

5.5 Choice of High-School Track and Teachers Recommendation

High-school track choice is the first crucial career decision in the Italian schooling system. Students together with their families are free to choose the track they like the most, with no constraints on grades and teachers' official track recommendation. There are three main types of high-school: academic, technical and vocational. 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 tracks. 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). Among students enrolled in vocational track, only 1.7% of the cohort graduating in 2016 enrolled to university, while the percentage raise up to 73.7% and 32.3% in the academic and technical track respectively. Interestingly, among students of the technical track the great majority enrolls in either STEM or economics degree: 62.5% vs. 52.4% of the academic track students (source: Miur).

I explore the impact of teacher bias on the track choice at the end of middle school, with a focus on the choice of the scientific academic track and on the vocational track. From a policy perspective the scientific academic path is particularly interesting since it easily opens up career opportunities in STEM related fields, while the vocational choice is highly correlated with low attendance to university degree. Table 9, Panel A, shows that girls are 9.4 percentage points less likely than boys to attend a scientific track and equally likely to attend a vocational track. Controlling for the standardized test score reduces half of the gap in the choice of scientific track, which is present also in track recommendations received from teachers (Panel B). However, I find a close to zero and insignificant effect of teacher bias on gender gap in scientific track choice (Panel A, columns 2-4) and in the recommendation of teachers toward a scientific track (Panel B, columns 2-4). The inclusion of controls at student and teacher level interacted with the pupil gender do not affect the point estimates of interest. In the questionnaire administered to teachers, I ask them why girls, compared to boys with the same math performance, are less likely to attend the scientific track: the reason 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.

Teacher stereotypes have stronger impact at the bottom of the ability distribution. Indeed, we can observe in columns 6 of Panel A that females, when assigned to a teacher with one standard deviation higher implicit bias, are more likely than their male classmates to attend vocational track 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. The subsequent two columns include characteristics of teachers and pupils and their interaction with the gender of the latter. Adding these controls does not change the coefficient of interest. However, what are female students choosing instead of vocational track when teachers have less bias? Girls assigned to teachers with higher implicit bias move from technical track to vocational, as shown in the Appendix Table A.9. Furthermore, in columns 4 and 8 of Table 1, we check whether reported gender bias is related with track recommendation or choice of students, but for both scientific and vocational track we find small and insignificant coefficients. Finally, appendix Table A.10 presents results from the heterogeneous analysis and, as expected, the impact of teacher bias has a stronger effect on the track choice of female students from disadvantaged background. The enrollment of females from the bottom tercile of the distribution increases of 4.3 percentage points for one standard deviation higher bias of the math teacher (which corresponds to a 15.8% increase with respect to the mean value for this group).

5.6 Additional Results and 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.11, I check whether math teachers bias has an impact on retention rate, but I do not find any significant impact, neither without nor with the inclusion of the standard controls at teacher and student level introduced in all specification throughout the paper. Furthermore, I also check that teacher implicit stereotypes does not differentially impact the probability of attending the standardized test score in grade 8 (Table A.11, 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. Additional checks on potential sample selection issues are addressed in the Appendix D.

All results exploit information on three cohorts of students. In the appendix Table A.12, I show the effect of the main specification presented in Table 5 for the three different cohort of students separately: reassuringly, results are not statistically different in the three cohorts, even if, since the number of observation decreases splitting the sample, estimates are noisier[30].

Table A.6 shows 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, indistinguishable from zero and not affected by inclusions of controls neither at Italian teacher level nor at pupil level. In the Appendix E, I delve deeper into the impact of Italian teacher on reading performance. Biased teachers activate stereotypes on female students only in male-typed domains.

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

Indeed, gender bias of Italian teachers have no statistically significant impact on reading performance of students, neither boys nor girls. 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 and 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 of the setting) and the type of task has an effect on the willingness to complete 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

In most OECD countries, women outnumber men in tertiary education, but they are by far a minority in highly paid fields as science, technology, engineering and math, especially when excluding teaching careers. The prospects are not optimistic considering that less than 5 percent of 15-years-old girls are planning to pursue a career in these fields, while the average share for boys is around 20 percentage points in OECD countries accordingly with 2015 PISA data. The cultural environment a person lives in has a strong impact on development of skills and educational choices. Indeed, this paper shows that gender gap in math performance can be partially explained by teacher implicit bias. Females, especially those from disadvantaged backgrounds, are lagging behind when assigned to teachers with higher implicit stereotypes (as measured by an Implicit Association Test). Males, the group not ability-stigmatized in terms of math performance, are not affected by teacher bias. Stereotypes foster low expectations about own ability and lead to under-performance. Biased teachers 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. These 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. Finally, 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

0.078 to 0.051standard deviation). 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.

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*. publisher not identified.

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.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics 117*(3), 871–915.

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. *The Quarterly Journal of Economics*.

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

Campa, P., A. Casarico, and P. Profeta (2010). *Gender culture and gender gap in employment*, Volume 57. Oxford University Press.

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*.

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.

Cooper, H. M. and T. L. Good (1983). *Pygmalion grows up: Studies in the expectation communication process*. Longman Publishing Group.

Cvencek, D., A. N. Meltzoff, and A. G. Greenwald (2011). Math–gender stereotypes in elementary school children. *Child development 82*(3), 766–779.

Dasgupta, N. and A. G. Greenwald (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology 81*(5), 800.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review 95*(2), 158–165.

Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry 52*(1), 173–182.

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

Fernández, R. (2007). Alfred marshall lecture: Women, work, and culture. *Journal of the European Economic Association*.

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.

Giavazzi, F., F. Schiantarelli, and M. Serafinelli (2013). Attitudes, policies, and work. *Journal of the European Economic Association 11*(6), 1256–1289.

Giustinelli, P. (2016). Group decision making with uncertain outcomes: Unpacking child–parent choice of the high school track. *International Economic Review 57*(2), 573–602.

Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*.

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., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74*(6), 1464.

Greenwald, A. G., B. A. Nosek, and M. R. Banaji (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology 85*(2), 197.

Greenwald, A. G., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji (2009). Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology 97*(1), 17.

Große, N. D. and G. Riener (2010). Explaining gender differences in competitiveness: gender-task stereotypes. *Jena economic research papers*.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Culture, gender, and math. *Science 320*(5880), 1164–1165.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *The Journal of Economic Perspectives 20*(2), 23–48.

Guryan, J. and K. K. Charles (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal 123*(572), F417–F432.

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

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).

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., 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. *The 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.

Rosenthal, R. and L. Jacobson (1968). Pygmalion in the classroom. *The urban review 3*(1), 16–20.

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

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

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

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

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

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

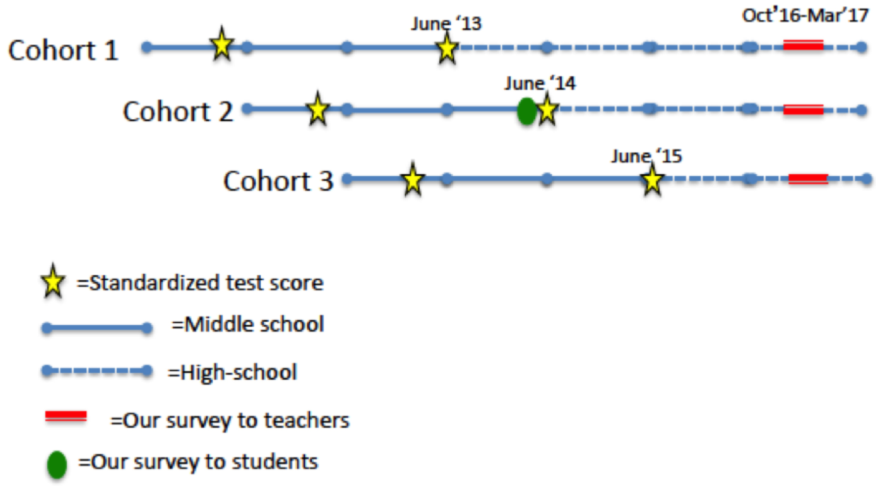

Figure 1: Timeline of main data available for students and teachers

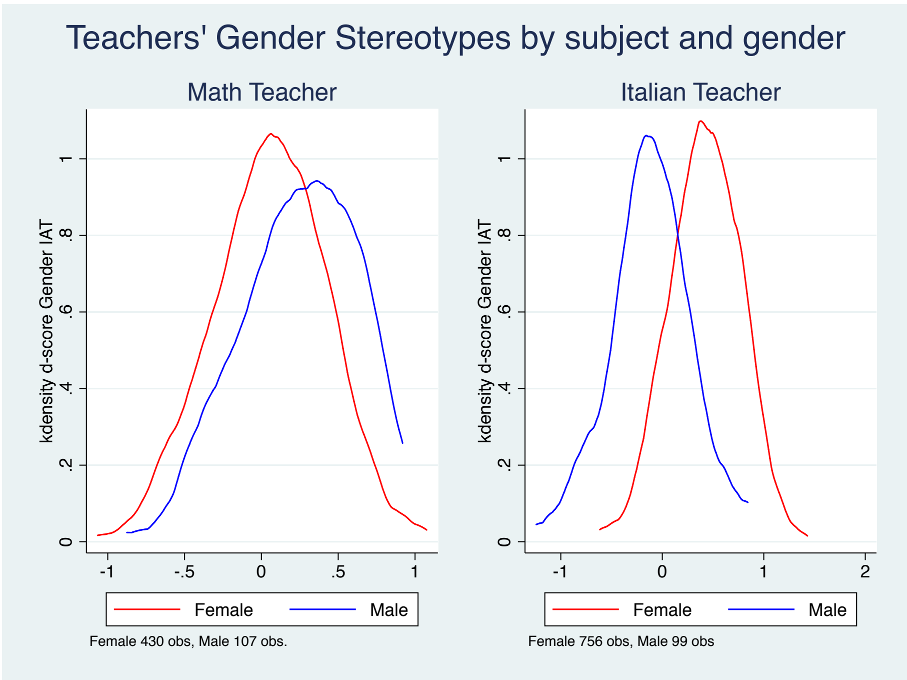

Figure 2: Teachers' Implicit Gender Bias (IAT measure) by gender and subject they teach

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	290	0.17	0.37	0.00	1.00
Gender Dif Innate Ability	280	1.51	0.76	1.00	3.00
Reason GenderGap: Interest for STEM	256	2.58	0.98	1.00	4.00
Reason GenderGap: Predisposition for STEM	241	2.12	1.03	1.00	5.00
Reason GenderGap: Low self-esteem	278	2.64	1.05	1.00	5.00
Reason GenderGap: Family support	278	3.14	1.08	1.00	5.00
Reason GenderGap: Cultural Stereotypes	279	2.15	1.16	1.00	5.00
Reported gender bias	206	0.00	1.01	-1.40	1.90
Boys better in Invalsi	233	0.20	0.40	0.00	1.00
Girls better in Invalsi	233	0.32	0.47	0.00	1.00
Gender Equal in Invalsi	233	0.48	0.50	0.00	1.00
Observations	301				

Notes: First-hand data from teachers' questionnaire. We restrict the sample to teachers matched to students and therefore used in the main analysis of this paper. The balance table with the difference between teachers' matched and not matched with students' data is presented in Table **??**. The main reason for not matching teachers with students is that they were not teaching in the school before 2016.

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

	Panel A: Independent variables (background teachers' characteristics)					
	Female	BornNorth	Age	HighMotherEdu	Children	Daughters
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:						
Raw IAT	-0.188**	-0.154**	0.016	-0.053	0.069	0.047
score	(0.083)	(0.064)	(0.060)	(0.060)	(0.145)	(0.075)
Obs.	301	301	301	301	301	301
R^2	0.347	0.348	0.327	0.337	0.330	0.331
	Panel B: Independent variables (education and teacher experience)					
	Ad.STEM	Laude	FullContract	Olympiad	JobSatisfy	Updates
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:						
Raw IAT	-0.092	-0.034	-0.049	0.059	0.054*	0.004
score	(0.076)	(0.075)	(0.153)	(0.087)	(0.032)	(0.037)
Obs.	301	301	301	301	301	301
R^2	0.332	0.326	0.327	0.311	0.336	0.325
	Panel C: Independent variables (beliefs)					
	WomenLFP	WVSCityBorn	WVSIndiv	InnateAbility	ExplicitBias	
	(1)	(2)	(3)	(4)	(5)	
Dep. Var.:						
Raw IAT	-0.499**	0.399*	0.007	0.016	0.047	
	(0.247)	(0.211)	(0.086)	(0.041)	(0.038)	
Obs.	286	261	301	301	301	
R^2	0.361	0.399	0.325	0.328	0.332	
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports OLS estimates of the correlation between math teachers' bias measured by IAT score and own teacher characteristics; the unit of observation is teacher t in school s. Standard errors are robust and clustered at school level in parentheses; the number of clusters is 90. School fixed effects are included in all regressions. The significance and magnitude of coefficients are not significantly impacted by the inclusion of FE. The variable "Female" indicates the gender of the teacher, "Born in the North " assumes value 1 if the teacher was born in the North of Italy, "HighMotherEdu" is a dummy which assumes value 1 is the mother of the teacher has at least a diploma,"Children" and "Daughters" are dummies which assumes value 1 if the teacher has children/daughters. The variable "Ad.STEM" assumes value 1 if the teacher has a degree in math, engineering and physics, "Laude" is a dummy which assumes value 1 if the degree was achieved with laude, "Full Contract" assumes value 1 is the teacher has tenure, "Olympiad" is 1 for teachers in charge of math Olympiad in the school, "JobSatisfy" is a categorical variable from 1 to 5 which captures self-reported job satisfaction of teachers, "Updates" captures whether teachers followed update courses in teaching during the academic year, "WomenLFP" is the labor force participation of women in the province of birth, "WVSCityBorn" is the WVS answer to the relative rights of men and women to paid jobs when the latter are scarce, "WVSIndiv" is the answer to the same question at individual level, "InnateAbility" regards the teacher belief about innate differences in math abilities between men and women,"ExplicitBias" is an index that summarizes explicit gender bias of teachers. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table 3: Summary Statistics of students by gender

	Males	Females	Diff.	se
Baseline characteristics				
Std Math grade 6	0.233	0.038	0.195***	(0.020)
Std Ita grade 6	0.085	0.218	-0.133***	(0.019)
Born in the North	0.849	0.854	-0.005	(0.007)
Born in the Center/South	0.027	0.030	-0.003	(0.003)
Immigrant	0.189	0.173	0.016	(0.008)
Second Gen. Immigrant	0.080	0.074	0.006	(0.006)
HighEduMother	0.456	0.453	0.003	(0.010)
Missing Edu Mother	0.212	0.211	0.002	(0.008)
High Occupation Father	0.169	0.174	-0.005	(0.008)
Medium Occupation Father	0.321	0.303	0.017	(0.010)
Missing Occupation Father	0.206	0.214	-0.008	(0.008)
Outcomes				
Std Math grade 8	0.194	-0.021	0.214***	(0.020)
Std Ita grade 8	-0.006	0.176	-0.182***	(0.020)
High-school Track: Scientific	0.304	0.208	0.096***	(0.010)
High-school Track: Classic	0.043	0.079	-0.036***	(0.005)
High-school Track: Other Academic	0.097	0.336	-0.239***	(0.009)
High-school Track: Technical Technological	0.311	0.067	0.244***	(0.008)
High-school Track: Technical Economic	0.113	0.163	-0.050***	(0.008)
High-school Track: Vocational	0.132	0.148	-0.015*	(0.008)
Track recommendation: Scientific	0.164	0.110	0.054***	(0.008)
Track recommendation: Vocational	0.362	0.298	0.064***	(0.011)
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)
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**	-0.037***	-0.035**
			(0.013)	(0.013)	(0.014)	(0.013)
Fem*Teacher Fem					-0.056	-0.046
					(0.037)	(0.037)
Fem*North Math Teacher					0.008	0.010
					(0.030)	(0.029)
Fem*Advanced STEM Teacher					-0.041	-0.037
					(0.031)	(0.031)
Fem*Reported Bias Teacher						-0.049***
						(0.017)
Std Math grade 6		0.723***	0.723***	0.697***	0.699***	0.698***
		(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Constant	0.198***	0.028***	0.028***	-0.112***	-0.112***	-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, age, children and daughthers, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, 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	-0.020	-0.054
	(0.022)	(0.015)	(0.015)	(0.033)	(0.107)	(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	-0.006
			(0.015)	(0.014)	(0.013)	(0.013)
Fem*Math Teacher Fem					-0.052	-0.045
					(0.040)	(0.040)
Math Teacher Fem					0.061	0.064
					(0.041)	(0.040)
Fem*North Math Teacher					0.013	0.014
					(0.031)	(0.030)
Math Teacher born North					0.027	0.025
					(0.035)	(0.035)
Fem*Advanced STEM Teacher					-0.031	-0.031
					(0.034)	(0.033)
Advanced STEM					0.026	0.026
					(0.034)	(0.035)
Fem*Reported Bias Teacher						-0.055***
						(0.016)
Reported Bias Teacher						0.016
						(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)
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, children and daughters, 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	-0.024	-0.020	0.033	-0.023	-0.050	-0.033
	(0.103)	(0.103)	(0.112)	(0.103)	(0.104)	(0.104)
Fem*Bias Teacher	-0.037***	-0.049**	-0.070***	-0.036**	-0.040**	-0.065**
	(0.014)	(0.021)	(0.027)	(0.015)	(0.016)	(0.031)
Fem*Bias T*HighEduM		0.022				
		(0.028)				
Fem*Bias T*Top tercile Math6			0.100***			
			(0.035)			
Fem*Bias T*Middle tercile Math6			0.011			
			(0.035)			
Fem*Bias T*Immigrant				-0.011		
				(0.038)		
Fem*Bias T*Extended School Day					0.012	
					(0.026)	
Fem*Bias T*Same Math Teacher						0.031
						(0.035)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	9309	9309	9309	9309	9309	9309
R^2	0.626	0.626	0.627	0.626	0.626	0.626

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on math standardized test score in grade 8 by observable characteristics of the student and by interaction time with teacher; the unit of observation is student i, in class c taught by teacher t in grade 8 of school s. Standard errors are robust and clustered at math teacher level in parentheses; the number of clusters is 301. The number of fixed effects (classes) is 548. The variable "Fem" indicates the gender of the student, "HighEduM" wether the mother has at least a diploma, "tercile Math6" is the tercile of standardized test score in math in grade 6 and "Immigrant" is a dummy equal to 1 if the student is not Italian citizen. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration and their interactions with the gender of the student. All columns include the interaction between math standardized test score in grade 6 and students' gender. Teacher controls include the interaction between students' gender and teacher gender, place of birth, age, children and daughthers, advanced STEM degree (as physics, math, engineering), leader of school math Olympics, degree with honor, update courses, type of contract and education of the teacher' mother. Regressions are all fully saturated even if not all interactions are shown in the table. *, ** and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table 8: Estimation of the effet of teachers' gender stereotypes on 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	0.273	-0.053*	-0.074	0.258
	(0.029)	(0.028)	(0.065)	(0.201)	(0.028)	(0.065)	(0.204)
Fem*Bias Teacher		-0.045**	-0.049**	-0.069**	-0.030	-0.033	-0.055*
		(0.021)	(0.022)	(0.031)	(0.021)	(0.023)	(0.032)
Constant	0.837***	0.808***	0.809***	0.799***	0.810***	0.820***	0.811***
	(0.015)	(0.015)	(0.048)	(0.047)	(0.015)	(0.048)	(0.046)
Std Test score math	No	Grade 6	Grade 6	Grade 6	Grade 8	Grade 8	Grade 8
Obs.	747	747	747	747	747	747	747
R^2	0.110	0.216	0.236	0.255	0.248	0.266	0.282
Panel B- Dependent Variable: Being good/mediocre at Italian (vs. being bad)							
Fem	0.052**	0.057**	0.045	0.315	0.047**	0.035	0.284
	(0.023)	(0.023)	(0.048)	(0.192)	(0.021)	(0.046)	(0.182)
Fem*Bias Teacher		0.038**	0.038**	0.028	0.038**	0.039**	0.030*
		(0.018)	(0.019)	(0.018)	(0.017)	(0.019)	(0.017)
Constant	0.916***	0.908***	0.937***	0.948***	0.917***	0.953***	0.964***
	(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	0.187	0.148	0.161	0.201
Panel C- Dependent Variable: Average own ability in other subjects							
Fem	0.035	0.019	0.021	-0.219	0.016	0.018	-0.223
	(0.027)	(0.029)	(0.062)	(0.223)	(0.028)	(0.062)	(0.225)
Fem*Bias Teacher		-0.014	-0.015	-0.027	-0.018	-0.020	-0.031
		(0.023)	(0.024)	(0.026)	(0.024)	(0.024)	(0.026)
Constant	1.672***	1.689***	1.674***	1.677***	1.687***	1.670***	1.673***
	(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	0.161	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, children and daughters, 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)
	Panel A- Dependent Variable: High-School Track Choice							
	Scientific Academic				**Vocational**			
Fem	-0.094***	-0.048***	0.171*	0.166*	0.014	-0.009	0.016	0.020
	(0.012)	(0.011)	(0.092)	(0.093)	(0.009)	(0.010)	(0.071)	(0.071)
Fem*Bias Teacher		0.009	0.001	0.001		0.023**	0.020**	0.020**
		(0.012)	(0.011)	(0.011)		(0.009)	(0.009)	(0.009)
Fem*Reported Bias				-0.004				0.005
				(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	0.236	0.236	0.119	0.190	0.211	0.211
	Panel B- Dependent Variable: Teachers' Recommendation							
	Scientific Academic				**Vocational**			
Fem	-0.045***	-0.019**	0.016	0.006	-0.059***	-0.110***	-0.059	-0.070
	(0.010)	(0.009)	(0.081)	(0.079)	(0.013)	(0.011)	(0.092)	(0.090)
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)	(0.010)
Fem*Reported Bias				-0.016				-0.009
				(0.012)				(0.012)
Std Math grade 6		0.126***	0.113***	0.113***		-0.246***	-0.217***	-0.218***
		(0.009)	(0.009)	(0.009)		(0.008)	(0.008)	(0.008)
Constant	0.156***	0.129***	0.059***	0.060***	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	0.252	0.150	0.362	0.391	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, children and daughters, 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.