



**UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO**

Department  
of Economics

## **WORKING PAPERS**

### **Gender stereotypes and children's performance in primary schools**

**Piera Bello, Annalisa Cristini, Federica Origo**

May 2024 - WP N. 25 Year 2024



**Working papers – Department of Economics  
n. 25**

**Gender stereotypes and children's performance  
in primary schools**



**UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO**

**Department  
of Economics**

**Piera Bello, Annalisa Cristini, Federica Origo**



---

**Università degli Studi di Bergamo  
2024**

**Gender stereotypes and children's performance in primary schools /  
Piera Bello, Annalisa Cristini, Federica Origo - Bergamo: Università degli  
Studi di Bergamo, 2024.**

**Working papers of Department of Economics, n. 25**

**ISSN: 2974-5586**

**DOI: [10.13122/WPEconomics\\_25](https://doi.org/10.13122/WPEconomics_25)**

**Il working paper è realizzato e rilasciato con licenza  
Attribution Share-Alike license (CC BY-NC-ND 4.0)**

**<https://creativecommons.org/licenses/by-nc-nd/4.0/>**

**La licenza prevede la possibilità di ridistribuire liberamente l'opera, a patto  
che venga citato il nome degli autori e che la distribuzione dei lavori derivati  
non abbia scopi commerciali.**



**Progetto grafico: Servizi Editoriali - Università degli Studi di Bergamo  
Università degli Studi di Bergamo  
via Salvecchio, 19  
24129 Bergamo  
Cod. Fiscale 80004350163  
P. IVA 01612800167**

**<https://aisberg.unibg.it/handle/10446/271950>**

# Gender stereotypes and children's performance in primary schools

Piera Bello, Annalisa Cristini, Federica Origo\*

\*University of Bergamo  
Department of Economics

## ABSTRACT

This paper investigates the impact of exposure to teacher stereotypes on student achievement in primary schools. We measure both implicit and explicit stereotypes using item-based questions available in the European Value Survey and the Gender-Science Implicit Association Test, respectively.

By exploiting the random assignment of students to teachers with different levels of stereotypes, we show that the gender gap in math performance, defined as the difference between boys' and girls' scores on standardized tests, significantly increases in classes with math teachers with stronger gender stereotypes. Additional evidence suggests that this result is driven by girls who benefit from having teachers with a girls-math attitude.

In line with previous research, we do not find any effect of teacher stereotypes on student outcomes in reading.

Keywords: gender stereotypes, education, human capital, implicit association test

JELs: J16, J24, I24

Acknowledgment: We are grateful to INVALSI, especially Patrizia Falzetti and Patrizia Giannantoni, the provincial working group for economic recovery and women's empowerment, as well as the seminar participants at the SABE-IAREP conference in 2023. The research is funded by the European Union -Next Generation EU, in the framework of the GRINS - Growing Resilient, Inclusive and Sustainable Project (GRINS PE00000018 – CUPE63C22002140007). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, and the European Union cannot be held responsible for them.

Corresponding author: Piera Bello [piera.bello@unibg.it](mailto:piera.bello@unibg.it)  
Università degli studi di Bergamo, Department of Economics,  
Via dei Caniana 2, Bergamo, Italy

Competing interests to disclose: none

Funding: none

## 1. INTRODUCTION

In recent decades, women's educational attainment and employment have both increased more rapidly than men's ones. However, large gender gaps persist in the labour market, especially in STEM occupations, where women have lower employment rates and wages than men.

One of the reasons given for these trends is the role of gender stereotypes and social norms. In particular, the preconception that girls are worse than boys in science and math and better than them in the arts and humanities may influence girls' and boys' perceived ability and school performance in different subjects. This is reflected in gender differences in aspirations and educational choices, with relatively few girls choosing technical high schools or STEM-related university degrees (especially in ICT and engineering), and subsequent gender gaps in occupations (Lavy and Sand, 2018; Guiso et al., 2008; Else-Quest et al., 2010).

In most OECD countries, the gender gaps in mathematics are typically small in the early years of primary school, but they increase significantly over time, peaking at the end of high school (Contini, Di Tommaso and Mendolia, 2017). This partly explains why most research on this topic focuses on secondary education (Niederle and Vesterlund, 2010; Rodriguez-Planas and Nollenberger, 2018; Burgess et al., 2022; Ellison and Swanson, 2023).

The influence of gender stereotypes in the early years of school is particularly important because the competence beliefs that children develop in these years are crucial predictors of their later choices and performance (Wigfield et al., 1997). Precisely because the early years of schooling lay the foundations for children's self-confidence and beliefs, it is crucial to understand the extent to which exposure to stereotypical primary school teachers affects the gender gaps in academic achievement. Evidence of a significant effect at this stage of education is essential for the design of timely interventions aimed at countering and disrupting existing stereotypes.

Primary school teachers, with whom children begin to build relationships, are an important channel through which gender roles are transmitted to children, either directly or indirectly (Alan et al. 2018). This is in addition to the influence of family and parents and, later peers who will play a major role in adolescence (Ertl et al, 2017; Nicoletti et al., 2022).

In this paper, we focus on children's academic achievements in second and fifth grade (grade II and V) and ask whether teachers' gender stereotypes influence the gender gap in students' math and reading scores.

Thanks to the support of the local school office, in the Spring 2022 we collected information on both implicit and explicit gender stereotypes from teachers in 40 different primary schools located in northern Italy (province of Bergamo). We were able to link each teacher to the class they taught in the 2021-22 school year.

Student performance is measured through standardized nationally administered and blindly scored tests in mathematics and reading that all students must take at the end of grade II and V of primary school. Given the timing of the national tests and of our teachers' survey, in this study we focus on pupils who were either grade II or V in 2021-22. The final sample is then composed of 78 teachers of either humanities or science matched with around 1500 students in grade II or V.

We measure both implicit and explicit gender stereotypes. Implicit gender stereotypes are elicited from the Gender-Science Implicit Association Test (IAT), i.e., a psychological test based on how quickly one associates female names with arts and humanities subjects and male names with science and math subjects (this method has also been used, for example by Carlana, 2019, and Brindusa et al., 2020). Explicit gender stereotypes are measured using item-based questions available in the European Values Survey. Respondents are asked, using a Likert scale, how strongly they agree with statements related to gender differences in different domains (for example: "Being a housewife allows a woman to fulfill herself as much as having a paid job", "When jobs are scarce, men should have

more right to a job than women”, “When a mother works for pay, the children suffer”, “If a woman earns more money than her husband, it's almost certain to cause problems”).

Given the characteristics of the Italian education system, in order to estimate the causal impact of teachers' gender biases on student's performance, our identification strategy exploits the random assignments of students to teachers with different levels of stereotyping.

The results of our analysis provide evidence that teachers' implicit biases play an important role in influencing students' test scores. In Grade II, classes assigned to teachers with one standard deviation higher IAT scores exhibit a 0.29 larger gender gap in math score. No significant effects are found in Grade V, although the signs are as expected. Interestingly, the educational background of teachers, especially education in pedagogical methods, is found to be important in mediating this effect.

A novel result of our analysis is that the effect is driven by girls taught by math teachers with unconventional implicit stereotypes, i.e. with a strong girls-math association. Specifically, in second grade, girls' math scores improve when they are taught by teachers with unconventional stereotypes and this effect dominates the decline in the scores when they are taught by teachers with conventional stereotypes (i.e., with a strong boy-math association).

In addition, we find evidence that mother's education is positively associated with children's reading scores, and the father's education with children's math scores, thus suggesting an intergenerational transmission of the bias.

In the growing recent literature on the effects of gender stereotypes on student achievement, the two studies closest to our paper are Alan et al. (2018) and Carlana (2019). They focus on primary schools in Turkey and middle schools in Italy, respectively.

We share with both studies an identification strategy based on random assignment of students to teachers with different levels of stereotyping.

While Carlana (2019) focuses on implicit gender stereotypes measured by the IAT, Alan et al. (2018) elicit explicit gender stereotypes through a battery of questions and summarize them with a score.

Estimates from Italy document that a change of one standard deviation on the IAT is associated with an increase of about one-third in the gender gap in mathematics in middle school. Interestingly, this is due to girls performing worse when assigned to a stereotyped teacher, while no effect is found for boys. The negative effect for girls is larger for those at the bottom of the distribution of the initial performance in mathematics and is partly due to girls' lower self-confidence in their mathematical abilities.

The analysis for Turkey shows that girls with teachers with traditional gender views (i.e., with higher explicit gender stereotypes) have lower performance in mathematics and verbal tests, and this effect is strengthened with longer exposure to the same teacher. On the contrary, no significant effects are found for boys.

Other studies (Lavy, 2008; Lavy and Sand 2019) develop alternative measures of teachers' gender discriminatory behaviour by computing the difference between the unblinded marks given by teachers to their students and the blinded marks of the same students on another test (the so-called “double-difference” method).

On the whole, most existing studies find a positive association between gender gaps in mathematics and teachers' gender stereotypes, especially in low-income countries and usually against girls. An important exception is Lavy (2008), who finds that stereotyped teachers discriminate against male students in all subjects. Avitzour et al. (2020) compute several indicators of both explicit and implicit stereotypes for a sample of about 100 primary school math teachers in Israel and find that implicit, but not explicit, gender stereotypes correlate with grading and assessment behaviour.

We contribute to the existing literature by providing new evidence on the impact of teachers' gender stereotypes on students' achievement at an early stage in their educational careers. Differently from Carlana (2019), we focus on primary school, which is a crucial stage for shaping students' beliefs and attitudes as well as their self-confidence. During these years, teachers play an important role in either challenging or reinforcing gender norms that may be transmitted to students by their parents. To the

best of our knowledge, Alan et al. (2018) is the only study that examines the effects of stereotypical teachers on students' achievement in primary schools. Our paper differs from this study in that we measure teachers' gender beliefs using the implicit association test (IAT), rather than relying solely on self-reported measures. This may be particularly relevant in the case of teachers who, due to their pedagogical role in the society, may be more likely than the rest of the population to provide socially desirable responses to explicit questions that elicit gender stereotypes, even if their answers do not correspond to their actual beliefs.

The remainder of the paper is organized as follows. In Section 2 we discuss the main institutional features of primary education in Italy. In Section 3 we present the data, the main variables of interests and some preliminary descriptive evidence. In Section 4 we discuss the empirical strategy, while the main econometric estimates, are reported in Section 5. In Section 6 we provide a number of robustness checks and further estimates on heterogeneous effects. The last Section concludes.

## 2. INSTITUTIONAL SETTING

Primary schools in Italy are attended by 6 to 11 years old students, thus for a duration of 5 years. The school year goes from middle September to early June, and children usually enroll in primary school in the school year in which they reach the age of six by the end of December. Early enrolment (that is one year earlier than the children born in the same cohort) is allowed for children born by the end April.<sup>1</sup> Throughout this period, students acquire skills in writing and reading, as well as introduction knowledge in subjects such as history, geography, mathematics, Italian grammar, science, English, music and sports activities.<sup>2</sup>

School hours may range from 24-27 to 30 hours per week, reaching 40 hours if pupils stay at school for the 2-hour lunch break (on a 5-day schedule). The latter corresponds to the full-time schedule, which is usually chosen by almost half of the families enrolling their child in the first year of a primary school, reaching 60% in most Northern regions and large cities.<sup>3</sup> Parents can formally choose their preferred time option within the same school, but since 30-hour and 40-hour classes can be formed only if schools have the necessary human resources (i.e., teachers) and facilities available (i.e., a canteen), the school week is usually school-specific.

More generally, parents can choose the primary school they prefer for their children, also on the basis of its distance from home or their workplace, or for the length of the school week. Schools accept applications, within the maximum limits of available slots, in accordance with criteria established by the School Council and published before the start of enrolment procedures. These criteria usually include whether the child lives in the neighborhood of the school, socio-economic status of the family and presence of older siblings in the same school. Since primary school is compulsory, all children must be enrolled in a primary school and this is facilitated by the coordination at the local level between schools and local administrations.

---

<sup>1</sup> Early enrolments (i.e., children aged under 6) are around 6% of total enrolment in the first year of primary schools in the school year 2021-22.

<sup>2</sup> Catholic religion is an elective subject. Furthermore, since September 2020, Citizenship education has been included as an additional and cross-subject topic in all grades of the Italian school system, with the aim to provide students with basic skills for an active citizenship, including basic knowledge of the Italian Constitution, sustainable development and digital skills.

<sup>3</sup> Regarding the other options, at the national level the second most frequent choice is 27 hours per week (around 30% of students), followed by 30 hours per week (20% of students). The short school week of 24 hours nowadays involves less than 3% of pupils in the first year of primary school.

Teaching activities are organized in classes, that is group of children that stay together during all the five years of primary school.<sup>4</sup> By law, the class size ranges from 15 to 26-27 students. Smaller classes are allowed in very small municipalities or in presence of children with disabilities or special educational needs.

The criteria for class formation are established by the school and are based on the principles of homogeneity between classes and similar heterogeneity within classes, according to dimensions such as gender, citizenship, disability and socioeconomic status of students. These criteria are usually made publicly available on the web site of the schools, in a document collecting all the Institute's regulations.<sup>5</sup>

Teachers in primary school are generalists, meaning that they do not specialize in a specific subject. A specific high-school pedagogical diploma<sup>6</sup> was required to become a primary school teacher until 2010; since then, teachers have to graduate in a specific 5-year university degree in primary education. The allocation of teachers to schools is administrated through a centralized system on the basis of seniority, which influences the likelihood of moving to the most desired school and location (Barbieri et al., 2011). Moreover, seniority affects the salary of teachers, which only depends on the length of the service and not on performance. The number of teachers per class varies with the length of the school week, provided that each teacher has to teach 22 hours per week. This implies that classes with short hours have usually one teacher for all subjects (except for English and Catholic religion, which are taught by specific teachers), while classes with 30 or 40 hours have at least two teachers. In this case, it is common that one teacher teaches Italian and other humanities-related subjects (such as history and geography), while the other teaches math and science. Normally the same teachers stay with the same class, teaching the same subjects, for the entire duration of primary school.

Students' performance in each subject is periodically evaluated by the corresponding teacher, usually on the basis of a 10-point scale. At the end of each term and at the end of the school year, pupils receive a personal assessment document reporting average marks by subject, as well as a brief descriptive report of their behaviour. External student-level administrative performance tests are managed by the National Institute for the Evaluation of the Italian Education System (INVALSI) and are administered in grade II and V. Therefore, during primary school, children take two tests. Both are paper tests and assess reading and math competencies.<sup>7</sup> Assignments are anonymously graded according to a detailed assessment criteria by an external teacher. Notice that the results of the national tests are not used to compute average students' marks, but they are returned to the school principals as useful information to improve teaching methods.

### 3. DATA AND DESCRIPTIVE STATISTICS

Data were collected as part of a larger post-pandemic project coordinated by the Province of Bergamo, whose main objective was to contribute to the recovery and resilience of the area by supporting women-related activities such as women's self-employment and gender-related issues in education. The Local Education Office (LEO), as well as the University of Bergamo, were also represented in the working group.

Figure A1 summarizes the main phases of the project from schools' involvement to the creation of the final dataset. In October 2021, in agreement with LEO, a letter was sent to all primary school principals in the province, introducing the research design and objectives and inviting them to a call. In that occasion we explained the main objectives of the study, providing basic definitions of both

---

<sup>4</sup> The only exception is due to students' (and their families) geographical mobility. Grade repetition is allowed in very exceptional circumstances and with parents' consent, such as in case of students with severe disabilities or new immigrant students who may benefit from staying longer in primary school.

<sup>5</sup> We inspected such documents for the primary schools in our sample. They report similar criteria for class formation.

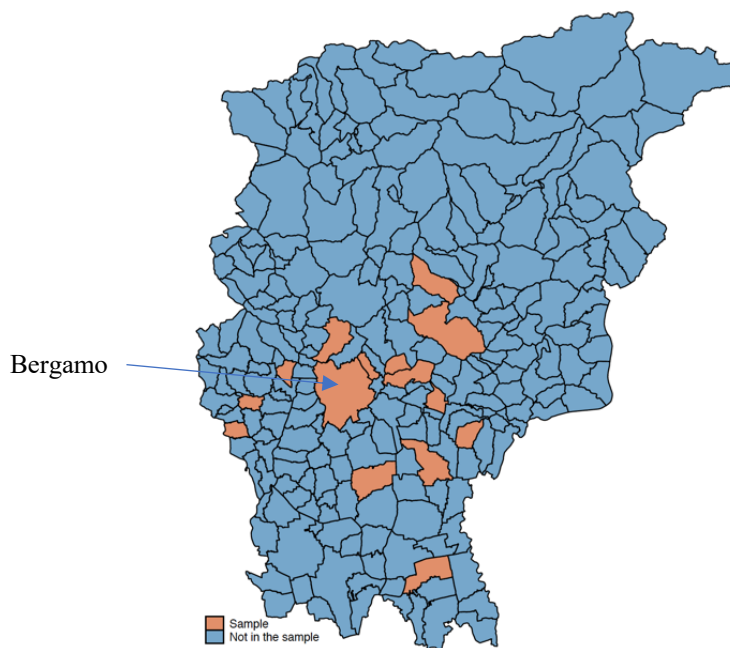
<sup>6</sup> The study plan of the psycho-pedagogical diploma includes up to 5 hours per week of psychology and pedagogy in the last three years of the course.

<sup>7</sup> Since 2018, students in grade V take also a standardized test to assess their English proficiency.

explicit and implicit stereotypes, and answered their questions. Following the meeting, which took place in March 2022, principals informed their teachers of the relevance and aim of the study and teachers were invited to participate. They were sent a link to take the IAT<sup>8</sup> and answer a questionnaire that included both explicit stereotypes questions taken from the European Value Survey, as well as personal information such as school, grade and subject taught, teaching experience, education, age, and gender (see Appendix). The questionnaires were collected automatically from April 30 to June 17, 2022. Overall, 196 teachers from 40 different primary schools completed the IAT and the survey. Notice that neither the school principals nor the teachers were informed about the detailed questions of the teachers' questionnaire and about the IAT in the previous phases of the projects. Furthermore, it is very unlikely that the teachers had the chance to practice with this test before our study. Hence, their answers should not have been influenced by previous knowledge about either the statements used to elicit explicit stereotypes or the IAT.

We link the teacher survey data and their IAT score with administrative information at the student level from INVALSI, which includes math and reading standardized scores for students enrolled in grade II and V in the school year 2021-2022.<sup>9</sup> The data also includes socio-demographic characteristics of the student, such as the date and place of birth, gender, citizenship and information on parents' education and occupation. The final sample is then made of around 1500 students in grade II and V matched with 78 unique teachers.

*Figure 1: Location of the schools included in our sample*



*Note: the map reports the municipalities of the province of Bergamo. Municipalities highlighted in red are those included in our analysis.*

Figure 1 shows the location in the province of the schools in our sample. The sampled schools are located in 16 different municipalities, including the province capital, mainly in the Center-South area of the province.<sup>10</sup> Moreover, in the appendix, Table A1 provides a balance test of the characteristics

<sup>8</sup> See Greenwald, McGhee, and Schwartz (2003) and Corno, Burns, and La Ferrara (2018) for more information on the functioning of the test.

<sup>9</sup> In that school year, all the teaching activities in primary schools were carried out in presence, including the standardized tests. The two tests were carried out in May, usually on different dates within the same week for students in the same class. Dates may slightly vary between classes and across schools.

<sup>10</sup> The Northern part is a mountains area, where there are smaller and more sparse schools.

of the students included in our dataset and those in all the schools in the province of Bergamo. The table shows that our sample is characterized by a higher level of immigrants, a lower share of students with parents with a graduate degree and a lower share of students that attended nursery or pre-school. However, the standardized difference for these variables is always below the threshold of 0.25 recommended by Imbens and Rubin (2005), with exception of the preschool and nursey variables. Moreover, no differences emerge for the average test scores.

Table 1 reports summary statistics on the teachers' characteristics, distinguishing between humanities teachers and science teachers. According to the table, 41% of humanities teachers and 30% of science teachers in our sample hold a university degree. Nearly 80% of them have attended a pedagogical high school. On average, they are over 40 age years old and 50% of them have more than 20 years of teaching experience.

Table 1: Summary Statistics on Teachers

Humanities Teachers				
	Mean	sd	min	max
Graduated	0.415	0.497	0	1
Pedagogical high school	0.811	0.395	0	1
Seniority >20	0.509	0.505	0	1
Age	47.170	9.585	25	63
IAT score	0.398	0.483	-0.903	1.372
Obs: 53				
Science Teachers				
	Mean	sd	min	max
Graduated	0.298	0.462	0	1
Pedagogical high school	0.894	0.312	0	1
Seniority >20	0.553	0.503	0	1
Age	48.362	9.352	26	63
IAT score	0.297	0.486	-0.903	1.059
Obs: 47				
Tot N Unique Teachers: 78				

Notes: The table shows descriptive statistics on teachers for the following variables: share of graduated teachers, share of teachers who have attended pedagogical high school, share of teachers with years of seniority larger than 20, average age, IAT score. Number of unique teachers 78. Note that the same teacher can teach in more than one class.

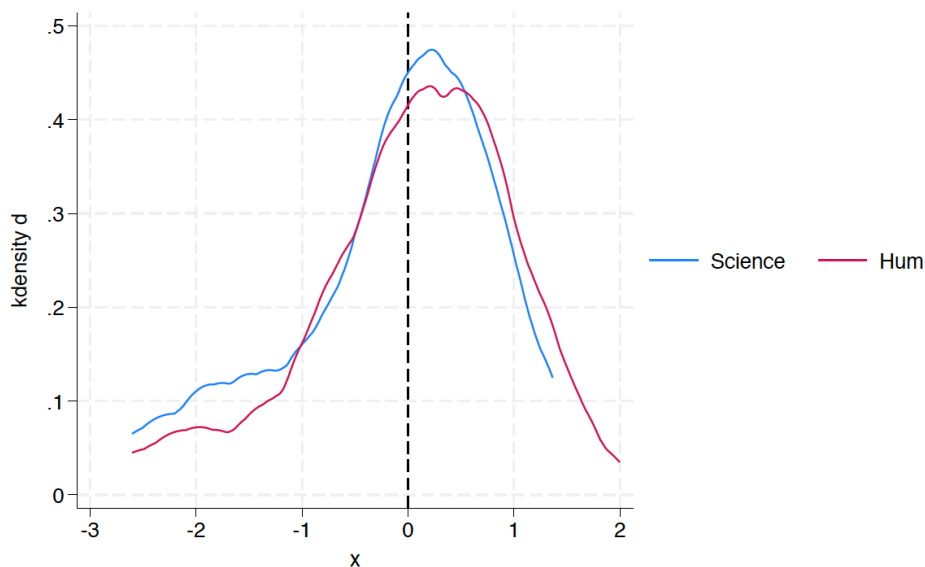
Regarding the IAT score, based on the estimates provided by psychological literature (Nosek et al, 2009) that suggests an average test score of 0.41, humanities teachers in our sample seem to be relatively close to the average: they have an IAT score of 0.40, while science teachers obtain a lower score than the average, namely 0.30. It is interesting to note that the difference in IAT test score between science and humanities teachers in our sample is significantly smaller compared to that reported in Carlana (2019) for a sample of middle schools in Italy. In that study, the average IAT test score for humanities teachers is 0.38 and that for math teachers is 0.09, while in our sample both groups of teachers seem to have a strong math-boys association. This points to the role that differences in educational background may have in explaining the observed variations in IAT score across teachers in the two different subjects. While in middle school a university degree in the specific subject of math or humanities is required to become a teacher in that field, the educational

requirements to become a primary school teacher are more homogenous across subjects, regardless of the specific subject being taught.

In the Appendix, in Figure A3, we also provide an analysis on the correlation between the IAT score and teacher's characteristics. We do not detect any significant correlation, except for a correlation between the IAT test score and pedagogical high school diploma.<sup>11</sup>

In the main analysis, to facilitate interpretation, we standardize the score to have a mean of 0 and a standard deviation of 1. The distribution of the standardized IAT test score is reported in Figure 2. In the appendix, in the Table A2, we also provide a t-test analysis to examine the differences in the baseline characteristics between science and humanities teachers for the II and V grade. No significant differences emerge.

Figure 2: Teachers' Implicit Gender Stereotypes by subject



Notes: This graph shows the distribution of Gender-Science IAT scores for science and humanities teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females.

Table 2 provides summary statistics for students included in our sample. According to the table, girls make up half of the sample, with 33% of students being foreigners. Additionally, 8% of students have mothers with a university degree and 5% have fathers with a university degree. Furthermore, 10% have attended nursery schools, and 78% have attended pre-school. Finally, in Table 3, we provide summary statistics on students test scores by gender. We standardized the test score by grade. Interestingly, the table reveals that while there are no differences in test score between girls and boys of grade II, a gender gap emerges in grade V. Girls outperform boys in humanities, while lagging behind in math. These findings suggest that the elementary school years are crucial for the development of students' skills and should be the focus of investigation when exploring the factors explaining the gender gap in STEM.

<sup>11</sup> The table shows that having a pedagogical high school diploma is associated with a one-point drop in the IAT test score.

Table 2: Summary Statistics on Students

	Mean	Sd
Girl	0.510	0.500
Nursery School	0.099	0.299
Pre-School	0.784	0.411
Foreigner	0.329	0.470
Schools in Bergamo	0.016	0.125
Graduated Mother	0.078	0.269
Graduated Father	0.046	0.210
Grade V	0.520	0.500
N:	1479	

Notes: The table shows descriptive statistics for students for the following variables: share of girls, share of kids who have attended nursery school or pre-school, share of foreigners, share of schools in the province of Bergamo, share of students with graduated mothers, share of students with graduated fathers, share of students enrolled in Grade V.

Table 3 Statistics on Student Test Scores by Gender (Outcome) (T-test analysis)

Panel 1: Grade II							
	Mean Boys	Mean Girls	t stat	Diff	p value	No. Males	No. Females
Math Test Score	0,04	-0,04	1,33	0,08	0,19	567,00	571,00
Reading Test Score	-0,02	0,02	-0,53	-0,03	0,60	564,00	572,00
Panel 2: Grade V							
	Mean Boys	Mean Girls	t stat	Diff	p value	No. Males	No. Females
Math Test Score	0,11	-0,10	3,66	0,22	0,000	538	578
Reading Test Score	-0,13	0,12	-4,39	-0,26	0,000	548	580

Notes: The table shows descriptive statistics by gender and, separately, for Grade II and Grade V for our outcome variables: the math test score and the reading test score

#### 4. EMPIRICAL STRATEGY

The identification strategy exploits the "as good as random" assignment of students to teachers with varying levels of implicit stereotypes, allowing us to provide causal evidence regarding the influence of gender stereotypes on gender gaps. This assumption stems from the features of the Italian school system described in Section 2: once parents have chosen the school for their children, the latter are allocated to classes – and hence associated to specific teachers - according to public criteria that cannot be controlled by parents, such as homogeneity between classes and similar heterogeneity within classes in terms of gender, citizenship, disability and socioeconomic status of students. It may be argued that parents can put pressure on school principals to put their children in classes with the best teachers. Even in this unlikely case, this should not be an issue in our empirical strategy, as long

as observable students' characteristics are not correlated with teachers' implicit gender bias, which should be unobservable also for the school principal.

Figure A2 in the appendix offers supportive evidence in this regard. It presents a regression coefficient plot, where the IAT score is regressed against a comprehensive range of student characteristics, such as family education and foreigner status. The plot demonstrates that there is no systematic correlation between students' baseline characteristics and teacher stereotypes, with the exception of attending pre-school and foreigners, which exhibits a weak correlation with TS.

We then proceed with our main analysis, following the approach adopted by Carlana (2019). We employ two identification strategies. First, we leverage the variation in gender gaps observed among students within the same classes. Formally, we estimate the following model:

$$y_{ic} = \alpha_0 + \alpha_1(Female_i * IAT_c) + \alpha_2 Female_i + \alpha_3 X_i + \alpha_4(Female_i * Z_c) + \eta_c + \varepsilon_{ic} \quad (1)$$

where  $y_{ic}$  is the outcome (i.e., math or humanities standardized test score) of student  $i$ .  $Female_i$  is a dummy variable that assumes value 1 if student  $i$  is a girl,  $IAT_c$  is the standardized value of the IAT score of teachers assigned to class  $c$  in the corresponding subject. Our variable of interest is the interaction between the latter two variables,  $Female_i * IAT_c$ , which measures the effect of teachers' stereotypes on the difference in test score between girls and boys.  $X_i$  is a rich set of students' characteristics (reported in the Table 2),  $\eta_c$  are class fixed effects<sup>12</sup>, which absorb differences in the teachers' characteristics across schools. Finally, in the full specification, we interact teachers' characteristics ( $Z_c$ ) with the gender dummy ( $Female_i$ ) to control for the fact that some teachers' characteristics might affect students' performance differently by gender. We estimate the model separately for humanities teachers and science teachers, respectively, for the reading test score and math test score.

The second identification strategy relies instead on the variation in the performance of students of the same gender across classes assigned to teachers with different level of stereotypes. This specification allows to explore whether any difference in the gender gap in classes assigned to teachers with stronger stereotypes eventually found with Model 1 is due to girls lagging behind, or to boys improving more, or a combination of these effects.

Formally:

$$y_{ics} = \beta_0 + \beta_1(Female_i * IAT_c) + \beta_2 Female_i + \beta_3 IAT_c + \beta_4 X_i + \beta_4(Female_i * Z_c) + \beta_5 Z_c + \eta_s + \varepsilon_{ics} \quad (2)$$

Where  $\eta_s$  are school fixed effect and all the other variables are defined as before.

As for the first model, we estimate the model using the reading and the math test score as dependent variables (and the IAT test score of the corresponding teachers among the regressors of interest).

## 5. RESULTS

Table 4 reports the results of Equation 1 for all the students included in our sample, while Table 5 distinguishes between students enrolled in grade II and those in grade V.

In Table 4, the estimates for the math test score are reported in the first two columns, while the estimates for the reading test score are presented in the third and four columns. Columns 1 and 3

regards the estimates of the model without the interaction between the girl dummy and the vector of teacher characteristics, while columns 2 and 4 report the results of the full specification.

The table documents several interesting results. First, it reveals the existence of gender gaps, with girls in math scoring -0.15 standard deviations lower than boys, while in reading, girls outperform boys with a score that is 0.16 standard deviations higher. Moreover, it highlights a significant and negative association between foreigner status and students' performance in both subjects. Additionally, it indicates a positive influence of parental education, with mothers' education having a stronger impact on the reading performance of the child, while fathers' one influencing their math performance.<sup>13</sup> This result is in line with Dossi et al (2021), whose analyses on the effect of parental preferences for boys and maternal gender attitudes on the math performance of girls confirm that the latter is influenced by preferences transmitted through the family.

Table 4: The impact of Teachers' Stereotypes, Model 1

VARIABLES	(1) Math	(2) Math	(3) Reading	(4) Reading
Girl	-0.149** (0.0609)	-0.352 (0.561)	0.159*** (0.0538)	-0.0638 (0.349)
Teacher Stereotypes*Girls	-0.0864 (0.0605)	-0.104* (0.0577)	0.0340 (0.0698)	0.0202 (0.0658)
Grade II	-0.0199 (0.185)	-0.0522 (0.192)	-0.267*** (0.0886)	-0.288*** (0.0999)
Foreigner	-0.510*** (0.0842)	-0.511*** (0.0846)	-0.644*** (0.0826)	-0.643*** (0.0824)
Graduated Mother	0.0707 (0.109)	0.0673 (0.110)	0.504*** (0.0877)	0.499*** (0.0875)
Graduated Father	0.484*** (0.0863)	0.481*** (0.0902)	0.0931 (0.162)	0.101 (0.159)
Nursery School	0.146 (0.0958)	0.141 (0.0950)	-0.128 (0.0855)	-0.128 (0.0854)
Pre-School	0.0754 (0.140)	0.0714 (0.141)	0.281*** (0.0790)	0.283*** (0.0790)
Graduated Teacher*Girls		0.00991 (0.204)		-0.0639 (0.136)
Seniority>20*Girls		-0.251 (0.218)		-0.140 (0.134)
Age Teacher*Girls		0.00794 (0.0108)		0.00707 (0.00561)
Pedagogical high school T*Girls		-0.0507 (0.143)		-0.0203 (0.167)
Observations	1,038	1,038	1,222	1,222
R-squared	0.261	0.263	0.221	0.222
Class FE	Yes	Yes	Yes	Yes
N Class	60	60	68	68

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 47 in columns 1 and 2, 53 in columns 3 and 4). Test score in Math/Reading. In all columns we include class FE.

<sup>13</sup> A similar result is also found in Carlana and Corno (2024).

Regarding the effect of teachers' stereotypes, although the coefficient of the interaction between the IAT variable and the girl dummy is almost never statistically significant, the sign and magnitude of the coefficient in columns 1 and 2 align with the hypothesis that a stronger math-boys association of the teachers negatively affects the relative math performance of girls as compared to boys.

Table 5: The impact of Teachers' Stereotypes, Model 1, Grade II

	(1)	(2)	(3)	(4)
	Math	Math	Math	Math
VARIABLES	Grade II	Grade II	Grade II	Grade II
Girl	-0.0407 (0.0807)	0.00775 (0.0831)	-0.0403 (0.694)	-0.328 (0.687)
Teacher Stereotypes*Girls	-0.300*** (0.0921)	-0.299** (0.106)	-0.293*** (0.0826)	-0.263** (0.102)
Observations	515	515	515	515
R-squared	0.218	0.289	0.291	0.305
N Classes	28	28	28	28
	(1)	(2)	(3)	(4)
	Reading	Reading	Reading	Reading
VARIABLES	Grade II	Grade II	Grade II	Grade II
Girl	0.0391 (0.0756)	0.0589 (0.0663)	-0.210 (0.412)	-0.544 (0.420)
Teacher Stereotypes*Girls	0.109 (0.0870)	0.103 (0.0715)	0.0978 (0.0675)	0.135** (0.0580)
Observations	622	622	622	622
R-squared	0.104	0.212	0.217	0.225
N Classes	34	34	34	34
Class FE	Yes	Yes	Yes	Yes
Students Controls	No	Yes	Yes	Yes
Teacher Controls*Girl	No	No	Yes	Yes
Student Controls*Girls	No	No	No	Yes

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in Panel 1, and 28 in Panel 2). Dependent variable: Test score in Math (Panel 1) and Reading (Panel 2), for students in Grade II. Column 1 includes only class\*teacher FE, while column 2 refers to a specification with also student controls. Finally, in column 3 and 4, we add, respectively, the interactions between teacher controls and the girl dummy, and the interactions between the latter and student controls.

Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Interestingly, when we estimate equation (1) by grade, in Table 5 and 6, we find stronger evidence of the detrimental impact of teachers' stereotypes on the math performance of girls, especially in grade II. Table 5 reports the results for the grade II, while Table 6 regards grade V.<sup>14</sup> In both tables, in Column 1, we provide the results of a specification without any control, in column 2, we add student controls, in column 3 we interact as before teacher characteristics with a dummy for girls, and finally, in the last column, we also interact student characteristics with the girl dummy. In Table 5, Panel 1, we observe that students in classes assigned to math teachers with a 1 standard deviation higher IAT

<sup>14</sup> Moreover, in Table A3 in the appendix, we report the coefficients of the full set of controls obtained in the specification with student controls and interactions between teacher controls and the gender dummy, for grade II and V, respectively.

score exhibit a 0.26-0.30 standard deviation larger gender gap in math performance in grade II. Notably, this coefficient is statistically significant at the 1 percent level in all columns. On the contrary, the coefficient of the interaction term of interest is not statistically significant for grade V (Table 6, Panel 1).

Finally, no effect is found on students' reading scores, in line with Carlana (2019), with exception of column 4, Table 5, Panel 2, where stronger boys-math stereotypes are positively associated with the relative performance of girls in reading. However, this effect is not precisely estimated in the other columns.

Table 6: The impact of Teachers' Stereotypes, Model 1, Grade V

	(1)	(2)	(3)	(4)
	Math	Math	Math	Math
VARIABLES	Grade V	Grade V	Grade V	Grade V
Girl	-0.256*** (0.0841)	-0.271*** (0.0808)	0.137 (0.621)	0.0944 (0.634)
Teacher Stereotypes*Girls	0.0260 (0.0534)	-0.0273 (0.0516)	-0.0217 (0.0617)	-0.0252 (0.0625)
Observations	523	523	523	523
R-squared	0.208	0.257	0.258	0.261
N Classes	32	32	32	32
	(1)	(2)	(3)	(4)
	Reading	Reading	Reading	Reading
VARIABLES	Grade V	Grade V	Grade V	Grade V
Girl	0.189** (0.0809)	0.265*** (0.0790)	-1.08e-05 (0.762)	-0.0755 (0.814)
Teacher Stereotypes*Girls	0.0630 (0.0904)	-0.0129 (0.0895)	-0.0329 (0.0798)	-0.0498 (0.0817)
Observations	600	600	600	600
R-squared	0.132	0.238	0.239	0.241
N Classes	34	34	34	34
Students Controls	No	Yes	Yes	Yes
Teacher Controls*Girl	No	No	Yes	Yes
Student Controls*Girls	No	No	No	Yes
Class FE	Yes	Yes	Yes	Yes

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 24 in Panel 1, and 25 in Panel 2). Dependent variable: Test score in Math (Panel 1) and Reading (Panel 2), for students in Grade II. Column 1 includes only class\*teacher FE, while column 2 refers to a specification with also student controls. Finally, in column 3 and 4, we add, respectively, the interactions between teacher controls and the girl dummy, and the interactions between the latter and student controls.

Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

In Table 7, we present the estimates of Equation 2, which allows us to examine the impact of teacher stereotypes on boys and girls separately. Specifically, the analysis compares, within the same school, the math and reading scores of boys and girls taught by teachers with different degrees of gender stereotypes. The aim is to disentangle whether the previous findings are driven by teachers with stronger math-boys associations positively benefiting boys' math performance or by the same teachers harming girls in this subject.

Table 7: The effect is due to boys or girls? Model 2

VARIABLES	(3)	(1)	(2)	(4)	(6)	(5)
	Math Grade II	Reading Grade II	Math Grade V	Reading Grade V	Math	Reading
Girl	0.00969 (0.0814)	0.0446 (0.0653)	-0.273*** (0.0801)	0.259*** (0.0787)	-0.0701 (0.410)	-0.0519 (0.347)
Teacher Stereotypes	0.159 (0.143)	0.113 (0.0710)	0.0443 (0.0710)	0.0716 (0.0564)	-0.0737 (0.0620)	-0.0311 (0.0520)
Teacher Stereotypes*Girls	-0.307*** (0.105)	0.106 (0.0725)	-0.0479 (0.0493)	-0.0188 (0.0889)	-0.0645 (0.0527)	-0.0321 (0.0692)
Grade II					-0.161* (0.0909)	0.0295 (0.0747)
Teacher Stereotypes*Grade II					0.254*** (0.0830)	0.164** (0.0746)
Girls*Grade II					0.273** (0.109)	-0.248** (0.103)
Teacher Stereotypes*Girls*Grade II					-0.256** (0.111)	0.106 (0.109)
Observations	515	622	523	600	1,038	1,222
R-squared	0.256	0.165	0.210	0.213	0.219	0.174
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Teachers Controls	Yes	Yes	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes	Yes	Yes
N Schools	12	14	13	13	15	16

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4, 47 in column 5, and 62 in column 6). Dependent variable: Test score in Math (columns 1,3,5), Reading (columns 2,4,6). We include school FE, student and teacher controls in all columns. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

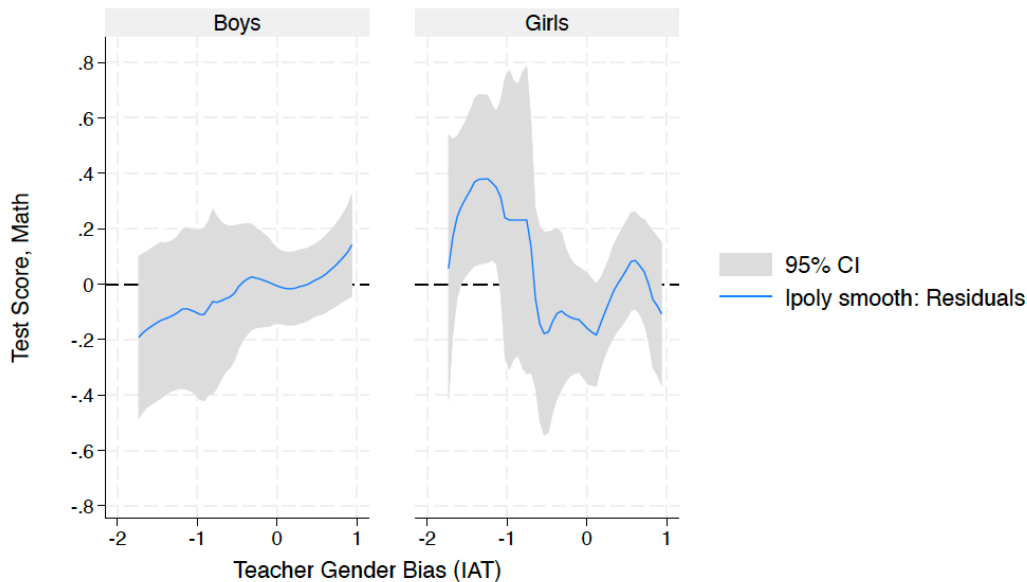
Notice that this specification allows the inclusion of school fixed effects but not of class fixed effects; hence, we add a rich set of controls for teacher and pupil characteristics to avoid omitted variable bias at the class level.

The results indicate that the coefficient of teacher stereotypes in column 1 is statistically significant and positive. This suggests that male students are not harmed from having teachers with a stronger association between boys and math, leading to improved math performance. However, the larger and negative coefficient of the interaction term suggests that the negative influence of stereotyped teachers on girls is much more pronounced than the benefits seen in boys.

In Figure 3, we better explore this point and plot the predicted test score estimated using Equation 2 over the distribution of the teachers' IAT test scores. According to the figure, no significant effect is found for boys along all the distribution and, if any, the math performance of boys slightly increases with the teacher's implicit bias. Having a highly stereotyped teacher does not seem to significantly influence girls' math performance either; however, girls significantly benefit from having a teacher with an "unconventional" gender stereotype, that is one with a stronger girls-math association. Overall, these results highlight that previous results are mainly driven by girls rather than by boys. Most importantly, they also suggest that to reduce the gender gap in math in primary schools it is

more crucial to have teachers with “unconventional” gender stereotypes rather than avoiding stereotyped teachers. According to our estimates, the gender gap in math should be in favor of girls in classes with such type of teacher.

Figure 3: The effect of teacher bias on student performance in math by gender, Grade II



Notes: The figure plots the residualized standardized math test score in grade II, after controlling for school fixed effects and teacher and student characteristics, along the IAT score of the teacher.

## 6. FURTHER ESTIMATES

### *Robustness checks*

We run several robustness checks to test the sensitivity of our results to the definition and measurement of gender stereotypes. First, we check that our results are robust to excluding teachers with a IAT score higher or lower than the 1st and 99th percentile of the IAT score distribution in our sample, respectively. The estimates provided in Table A4 (Panel 1) confirm that our results are not driven by outliers.

Second, we replicate our baseline analysis using a dummy variable for teachers with a boy-math association instead of the continuous IAT score variable. Hence, our variable of interest is the interaction between a IAT dummy equal to 1 for those teachers with a IAT score greater than zero and the girl dummy. The results are provided in Table A4 (Panel 2) and confirm the robustness of our results. The magnitude of the effect appears to be even larger compared to Table 2. More specifically, having a stereotyped teacher reduces the test score in math more for girls than for boys (by around 0.44 standard deviation of the math test score).

Finally, it may be argued that the IAT test score is highly correlated with explicit stereotypes, which are easier to elicit in official surveys. In the Introduction, we have already pointed out that measures of explicit stereotypes may be highly influenced by social desirability in self-reported opinions on specific statements related to gender differences, especially in the case of teachers. Our data confirm that this may be the case: very few teachers declared to agree or strongly agree with most of the statements that we used in the survey to elicit gender stereotypes. The only statement for which we find some variability in self-reported answers is: “Being a housewife allows a woman to fulfill herself as much as having a paid job”. We then generate a dummy variable equal to 1 if the teacher states

that she agrees or strongly agrees with the statement (the dummy has mean 0.25 and standard deviation 0.43).<sup>15</sup> In Figure A4, we also report the binscatter of the correlation between the self-reported level of agreement to this statement and the IAT test score of the teachers. The correlation is equal to 0.05 and is not statistically significant. Finally, we test if our main results survive when we control for explicit gender stereotypes. In particular, we re-estimate Equation 2 including the dummy for explicit gender stereotypes and its interaction with the Girl dummy as additional regressors. The results are reported in Table A6. We find a weak association between explicit gender stereotypes and gender gap in math in both Grade II and Grade V. Quite interestingly, our variable of interest (Teacher stereotypes\*girl) is still negative and statistically significant (see estimates in column 1 of Table A6). This confirms that indicators of explicit stereotypes are not always good proxies for implicit stereotypes, and collecting data on the latter is important to investigate their effects on students' performance.

### ***Heterogeneous effects: the role of teacher's education background***

Since descriptive evidence shown in Table A2 suggests that teacher's education background is might be a determinant of gender stereotypes, we investigate whether it also influences how gender stereotypes affects the gender gap in students' performance. Specifically, we expect that teachers with a strong pedagogical background may be abler to adapt their teaching methods to the needs of different students, using a student-centered approach rather than a teacher-centered approach.<sup>16</sup> This should contribute to reduce the gender gap in students' performance.

In order to test this hypothesis, we interact our variable of interest "Teacher Stereotypes\*Girls" with two different variables measuring teachers' education background: in a first specification, we use a dummy variable for teachers who attended a pedagogical high-school; alternatively, we use a dummy variable for graduated teachers. As discussed in the Section on the institutional setting, since 2010 a specific 5-year degree in primary education is required to teach in primary schools. We estimate an Equation similar to Equation 2. The results are reported in Tables 8 and 9, respectively. Interestingly, Table 8 shows that the attendance of a pedagogical high-school slightly reduces the effect of implicit gender stereotypes on the gender gap in math in Grade II (the coefficient of the variable named "Pedagogical HS \*Teacher Stereotypes\*Girls" is positive and statistically significant at 10 percent level in column 1), while no evidence for a different effect between graduated and non-graduated teachers is found (the coefficient of the variable labelled "Graduated T\*Teacher Stereotypes\*Girls" is not statistically significant in column 1 of Table 9).<sup>17</sup> However, the university degree seems to play a protective role in grade V: the effect of gender stereotypes on the gender gap in math is significantly lower in classes where the math teacher holds a university degree in primary education, while no significant heterogeneous effects are found in grade V for the type of high school (compare the estimated coefficient for the relevant triple interaction terms in column 3 of Table 8 and Table 9)

---

<sup>15</sup> For all the other statements on explicit stereotypes, less than 10% of the teachers state that they agree with the statement. The full list of statements is in Table A8.

<sup>16</sup> This is a crucial issue in Italy, which stands out as the OECD country with the lowest share of teachers adopting teaching practices centered on students (OECD, 2009).

<sup>17</sup> On the contrary, the coefficient of our variable of interest is not statistically significant in column 3, which regards the performance of students in grade V in math. Notice that the coefficient of our variable is instead negative and slightly significant in reading score in grade V (column 4) again suggesting that teachers with a pedagogical HS training can help smoothing gender differences in math and reading.

Table 8: The impact of the teacher's background (pedagogical high school) (Model 2)

VARIABLES	(1)	(2)	(3)	(4)
	Math	Reading	Math	Reading
	II	II	V	V
Girl	0.0624 (0.0531)	0.0996 (0.0719)	-0.114** (0.0490)	-0.477** (0.213)
Teacher Stereotypes	0.418 (0.575)	-0.323*** (0.0632)	0.299 (0.467)	-0.122* (0.0706)
Teacher Stereotypes*Girls	-0.516*** (0.0771)	0.112 (0.0723)	0.127 (0.174)	0.729*** (0.0963)
Pedagogical high school Teacher	-0.372* (0.206)	0.0499 (0.0719)	0.260 (0.238)	0.193* (0.111)
Pedagogical HS*Teacher Stereotypes	-0.267 (0.707)	0.520*** (0.0723)	-0.252 (0.476)	0.229*** (0.0681)
Pedagogical HS*Girls	-0.0628 (0.107)	-0.0674 (0.107)	-0.179* (0.0972)	0.727*** (0.226)
Pedagogical HS*Teacher Stereotypes*Girl	0.222* (0.115)	-0.0170 (0.120)	-0.181 (0.200)	-0.826*** (0.113)
Observations	515	622	523	600
R-squared	0.257	0.174	0.211	0.229
School FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
N Schools	12	14	13	13

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Overall, these findings highlight the importance of perspective teachers' exposure to pedagogical methods as a potential channel to mitigate the impact of their stereotypes on students' performance, especially in the first years of primary school. In a policy perspective, these results also provide valuable insights into potential remedies that can be implemented to mitigate the negative effects of implicit gender stereotypes on the gender gap in math, such as training programs focused on pedagogical knowledge in the teachers' subject field. Evidence from the 2013 OECD Teaching and Learning International Survey (TALIS), the largest international survey of teachers, confirms that this type of training, together with professional development that focus on subject-specific contents, is perceived as particularly helpful also by the teachers themselves (OECD, 2014).

We carry out further analyses to investigate the existence of additional heterogeneous effects by other teacher and student characteristics. Specifically, we analysed the effect of years of exposure to stereotyped teachers (Table 10) for classes in grade V. We interact our variable of interest with a dummy for those teachers who have been teaching in the same class for more than two years. Interestingly, we find that the coefficient of the triple interaction is significant and negative at 10

percent level in column 1. This suggests that in those classes with a longer exposure to stereotyped teachers, teacher gender stereotypes slightly affect the gender gap in math also for grade V. We also examine the role of both mother's and father's education. However, the coefficients of the variables of interest are not statistically significant; hence, we could not detect heterogeneous effects by parental education.<sup>18</sup>

Table 9: The impact of the teacher's education level (graduated vs non graduated) (Model 2)

VARIABLES	(1)	(2)	(3)	(4)
	Math 2	Reading 2	Math V	Reading V
Girl	0.00530 (0.0945)	-0.00320 (0.0888)	-0.309** (0.111)	0.325*** (0.0991)
Teacher Stereotypes	0.337*** (0.104)	0.183** (0.0736)	0.0889 (0.0860)	0.0554 (0.0612)
Teacher Stereotypes*Girls	-0.265*** (0.0930)	0.136 (0.109)	-0.0950 (0.0577)	0.0311 (0.0924)
Graduated Teachers	-0.398*** (0.0718)	-0.174 (0.116)	-0.221 (0.242)	0.294** (0.133)
Graduated T*Teacher Stereotypes	-0.546*** (0.143)	-0.200 (0.156)	-0.209* (0.120)	-0.00308 (0.0666)
Graduated T*Girls	-0.0605 (0.132)	0.103 (0.125)	0.117 (0.124)	-0.133 (0.157)
Graduated T*Teacher Stereotypes*Girl	-0.214 (0.217)	-0.0551 (0.141)	0.224** (0.0992)	-0.266 (0.169)
Observations	515	622	523	600
R-squared	0.269	0.167	0.213	0.220
School FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
N Schools	12	14	13	13

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

<sup>18</sup> Estimates are available upon request.

Table 10: The impact of the length of exposure (exposure same teacher>2) (Model 2)

VARIABLES	(3)	(4)
	Math V	Reading V
Girl	-0.152*** (0.0400)	0.198* (0.107)
Teacher Stereotypes	0.640 (0.580)	0.0810 (0.113)
Teacher Stereotypes*Girls	0.0796 (0.0502)	-0.0573 (0.128)
More than 2 years same class	0.158 (0.115)	0.0491 (0.112)
More than 2 years same class*Teacher Stereotypes	-0.550 (0.535)	0.00419 (0.120)
More than 2 years same class*Girls	-0.159 (0.107)	0.0765 (0.146)
More than 2 years same*Teacher Stereotypes*Girl	-0.155* (0.0809)	0.0386 (0.154)
Observations	523	600
R-squared	0.213	0.214
School FE	Yes	Yes
Students Controls	Yes	Yes
Teacher Controls	Yes	Yes
N Schools	13	13

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 24 in column 1, 25 in column 2). Dependent variable: Test score in Math/Reading. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher. Grade V.

## 7. DISCUSSION AND CONCLUSIONS

### *Discussion*

The different role that the teacher's implicit gender stereotype plays in grade II compared to grade V can be ascribed to at least three possible and not mutually exclusive explanations.

The first mechanism is related to the changing role of parents. As shown in Table A3, which reports the coefficients of the full set of controls used in the previous specification (Tables 5 and 6, column 3), while in grade II mother's education affects child's score in both math and reading, in grade V mother's education affects the score only in reading, and father's education affects the score in math. Results for grade V are then consistent with parents' gender stereotypes, which may partly prevail on teachers' gender stereotypes in influencing student's performance. In the previous Section, we tried to investigate this dynamic by interacting parental education with the teacher stereotypes variable. Although we did not find significant results, we believe that this mechanism deserves further investigation.

A second factor, coherent with the previous explanation, rests on the teacher-student relationships. The education and developmental psychology literature finds that the closeness in teacher-student relationships diminishes as children advance through school (Jerome et al., 2009; O'Connor and McCartney, 2007; Maldonado-Carreño and Votruba-Drzal, 2011) and that it is mainly girls' math performance to be affected by the teacher-child relationship (McCormick et al., 2015). The protective effect for girls of being matched with a teacher with unconventional bias that we find for grade II but not for grade V is coherent with this evidence.

Finally, it is possible that the COVID-19 pandemic and the consequent turnover of teachers might have affected students in grades II and V differently.<sup>19</sup> First, school closure during the harshest periods of the pandemic, especially during the lockdown in spring 2020 and the second wave at the beginning of 2021, has reduced direct exposure to stereotyped teachers for students in Grade V. Second, interruption of teachers' careers because of COVID, by shortening the length of exposure to the same teachers, might have weakened the role of this channel for students in grade V. Unfortunately, we do not have detailed information to delve into the role of teachers' turnover, but suggestive evidence in Table 10 shows that classes in grade V with longer exposure to stereotyped teachers also register higher gender differences in math.

## ***Conclusion***

The underrepresentation of girls and women in science, technology, engineering, and mathematics (STEM) fields is a long-lasting concern for policy makers. The persistent nature of this gap, coupled with its correlation with the level of national gender equality across countries, points to the role played by gender norms and stereotypes in explaining this imbalance.

In this study, we delve into the role played by teachers and investigate whether teachers' gender stereotypes influence gender gaps in student performance by exploiting random allocation of students to teachers in primary schools in Italy.

Our study shows that gender gap in student performance is significantly influenced by the implicit gender biases held by teachers. Classes where teachers have more stereotypical attitudes exhibit a more pronounced gender gap in math performance. Conversely, when girls are taught by teachers with less traditional gender perspectives, their math performance tends to be higher. Interestingly, neither girls nor boys seem to be affected by gender stereotyped humanities teachers.

Two main channels may actually explain the positive effects of teachers with unconventional gender stereotypes on girls' math performance. First, in line with models documenting the role of stereotypes on the formation of beliefs and psychological attitude (Jouini et al., 2018), teachers with a strong girls-math association may support girls' self-confidence and perceived ability in math, with subsequent effects on their performance. An alternative explanation may rely on the adoption of teaching practices that allow girls to learn math more effectively. In this respect, evidence from a small-scale Randomized Control Trial carried out in primary schools in Italy shows that more active teaching methods (involving intense peer interactions, learning from mistakes and problem solving) significantly increase girls' math performance, with no impact on boys (Di Tommaso et al., 2021). Our data do not allow to disentangle between these two mechanisms, which are left for future research.

These findings underscore the critical role of teachers in influencing gender gaps in student performance, particularly in math, and emphasize the importance of addressing implicit gender biases in educational settings with targeted policy interventions. These interventions could include enhancing teachers' understanding of cultural stereotypes and implementing measures to counteract them through specific training measures.

Additionally, our results suggest that the educational background of teachers can serve as a protective factor against the perpetuation of gender disparities in academic achievement. From a policy

---

<sup>19</sup> We thank one of the referees for suggesting this possibility.

perspective, teacher training and education can be leveraged to address and counteract the influence of biases on student outcomes. By fostering an inclusive and unbiased pedagogical approach, educators can create a more equitable learning environment, promoting better performance and opportunities for all students, regardless of their gender.

This research contributes to the understanding of gender gaps in education and labor markets with the final scope to promote fairness and equality in education and opens avenues for targeted interventions to improve educational outcomes and close the gender gap in math, which encompass not solely the awareness-raising of teachers regarding their inherent biases but also training courses to guarantee equitable treatment of all students.

## REFERENCES:

1. Alan, Sule, Ertac, Seda, Mumcu, Ipek. (2018). "Gender Stereotypes in the Classroom and Effects on Achievement." *Review of Economics and Statistics*, v. 100, iss. 5, pp. 876-90.
2. Avitzour, Eliana, Adi Choen, Daphna Joel, Victor Lavy. (2021). "On The Origins Of Gender-Biased Behavior: The Role Of Explicit And Implicit Stereotypes." NBER Working Paper 27818.
3. Brindusa, Anghel, Rodriguez-Planas, Nuria, Sanz-de-Galdeano, Anna. (2020). "Is the Math Gender Gap Associated with Gender Equality? Only in Low-Income Countries." *Economics of Education Review*, December 2020, v. 79.
4. Burgess, S., Hauberg, D.S., Rangvid, B., Sievertsen, H. (2022). "The Importance of External Assessments: High School Math and Gender Gaps in STEM Degrees." *Economics of Education Review*, v. 88.
5. Carlana, M. (2019). "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics*, 1163-1224.
6. Carlana, M., Corno, L. (2024) "Thinking about Parents: Gender and Field of Study" *AEA Papers and Proceedings*. Vol. 114, No. : Pages 254-258.
7. Contini, D., Di Tommaso, M.L., Mendolia, S. (2017). "The Gender Gap in Mathematics Achievement: Evidence from Italian Data." *Economics of Education Review*, 58, 32–42.
8. Corno, Lucia, Justine Burns, and Eliana La Ferrara. "Interaction, Stereotypes and Performance. Evidence from South Africa." BREAD Working Paper no. 549, 2018.
9. Di Tommaso, M.L., Contini, D., De Rosa, D., Ferrara, F., Piazzalunga, D., Robutti, O. (2021). "Tackling the Gender Gap in Math with Active Learning Methodologies." IZA DP 14572.
10. Dossi, Gaia, Figlio, David, Giuliano, Paola, Paola Sapienza. (2021). "Born in the Family: Preferences for Boys and the Gender Gap in Math." *Journal of Economic Behavior and Organization*, v. 183, pp. 175-88.
11. Ellison, G., Swanson, A. (2023). "Dynamics of the Gender Gap in High Math Achievement." *Journal of Human Resources*, v. 58, iss. 5, pp. 1679-1711.
12. Ertl, Bernhard, Silke Luttenberger, Manuela Paechter. "The Impact of Gender Stereotypes on the Self-Concept of Female Students in STEM Subjects with an Under-Representation of Females." *Frontiers in Psychology*, 8 (2017), 703.
13. Greenwald, Anthony G., Brian A. Nosek, Mahzarin R. Banaji. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology*, 85 (2003), 197.
14. Guiso, Luigi, Ferdinando Monte, Paola Sapienza, Luigi Zingales. "Culture, Gender, and Math." *Science*, 320 (2008), 1164–1165.
15. Jerome, E., Hamre, B., & Pianta, R. C. (2009). Teacher–child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development*, 18, 915–945.
16. Hyde, Janet S., Elizabeth Fennema, Susan J. Lamon. "Gender Differences in Mathematics Performance: A Meta-Analysis." *Psychological Bulletin*, 107 (1990), 139.
17. Jouini, E., Karehnke, P., Napp, C. (2018). "Stereotypes, Underconfidence and Decision-Making with an Application to Gender and Math." *Journal of Economic Behavior & Organization*, v 148, p. 34-45.
18. Lavy, Victor. (2008). "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics*. DOI link.
19. Lavy, Victor, Edith Sand. (2018). "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases." *Journal of Public Economics*. DOI link.
20. Maldonado-Carreño, C., and Votruba-Drzal, E. (2011). Teacher–child relationships and the development of academic and behavioral skills during elementary school: A within- and between-child analysis. *Child development*, 82(2), 601-616.

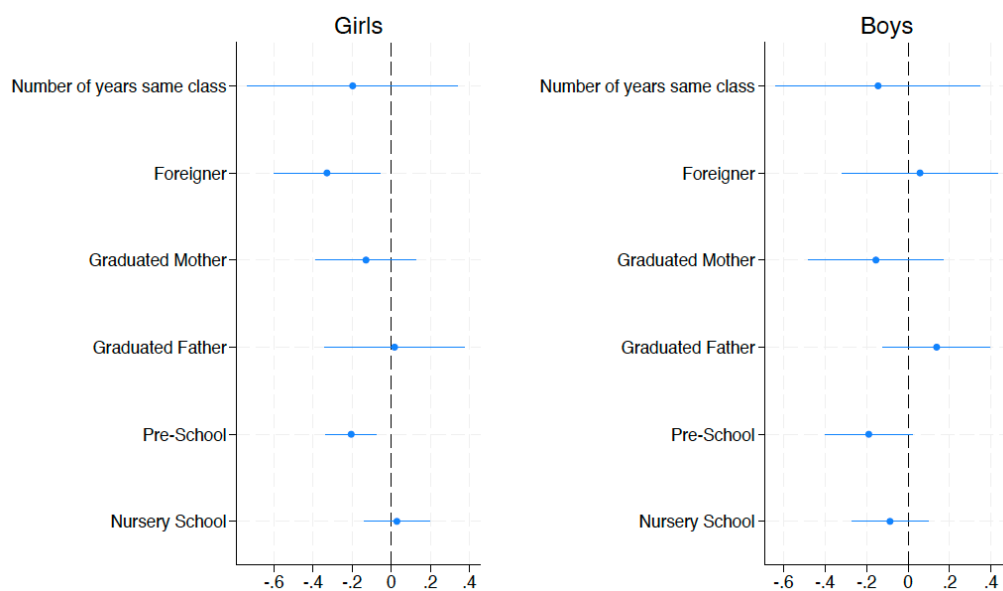
21. McCormick, Meghan P., and Erin E. O'Connor. "Teacher–child relationship quality and academic achievement in elementary school: Does gender matter?." *Journal of Educational Psychology* 107.2 (2015): 502.
22. Nicoletti, Cheti, Almudena Sevilla, Valentina Tonei. "Gender Stereotypes in the Family." (2022).
23. Niederle, M., Vesterlund, L. (2010). "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives*, v. 24, iss. 2, pp. 129-44.
24. Nosek, Brian A., and Frederick L. Smyth. "Implicit social cognitions predict sex differences in math engagement and achievement." *American Educational Research Journal* 48.5 (2011): 1125-1156.
25. OECD (2009). "Creating Effective Teaching and Learning Environments First Results from TALIS". OECD Publishing, Paris.
26. OECD (2014), "Talis 2013 Results: An International Perspective on Teaching and Learning". OECD Publishing, Paris.
27. OECD (2019). "Why Don't More Girls Choose to Pursue a Science Career? PISA in Focus #93." OECD Publishing, Paris.
28. Rodriguez-Planas, N., Nollenberger, N. (2018). "Let the Girls Learn! It Is Not Only about Math . . . It's about Gender Social Norms." *Economics of Education Review*, v. 62, pp. 230-53.
29. Wigfield, Allan, et al. "Change in Children's Competence Beliefs and Subjective Task Values across the Elementary School Years: A 3-Year Study." *Journal of Educational Psychology*, 89.3 (1997), 451.

## APPENDIX

*Figure A1: Time line Project*

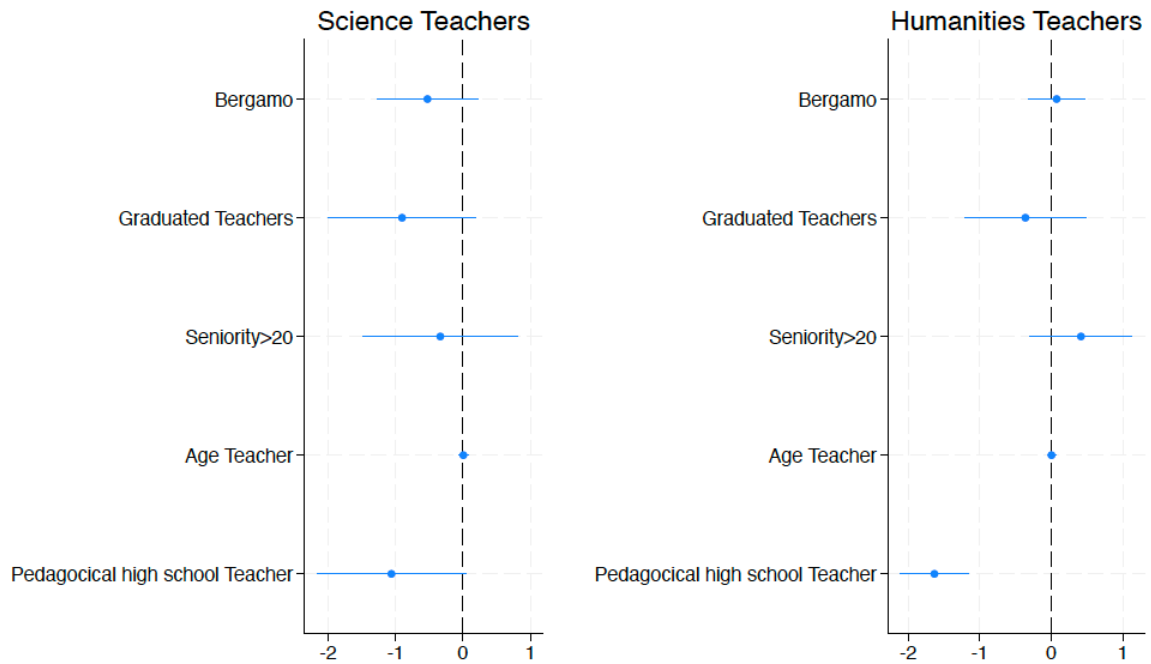
	2021		2022				N schools	N. unique teachers
	October	March	April	May	June	October		
Letter on the project sent by Local Education Office to all primary schools in the province	■						283	3628
Online meeting with school principals		■						
Schools officially joining the project*			■	■	■		40	196
Online Teachers survey				■	■			
National standardized tests (Invalsi)						■		
Release of Invalsi microdata and merge with teachers' survey						■	31	78

*Figure A2: Correlation between students' characteristics and teachers' IAT test scores*



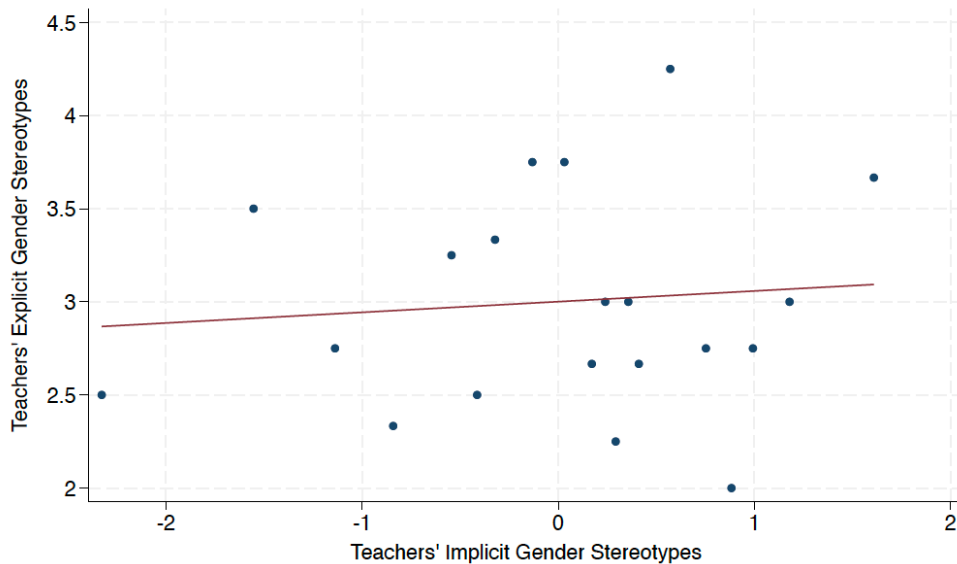
*Notes: The figure reports OLS estimates of the correlation between science teachers' IAT score and students' characteristics. The estimates come from a model with school FE and standard errors clustered at the school level.*

Figure A3: Correlation between Gender IAT Score and Teachers characteristics by Field



Notes: The figure reports OLS estimates of the correlation between teachers' IAT score and own characteristics. The estimates come from a model with school FE and standard errors clustered at the school level.

Figure A4: Correlation between Explicit and implicit Stereotypes



Notes: The binscatter shows the correlation between the explicit stereotype variable and the IAT Test score. In order to measure the level of teacher's explicit stereotypes we use the self-reported level of agreement with this statement "Being a housewife allows a woman to fulfil herself as much as having a job". A higher level of agreement with this statement translates in a higher level of explicit stereotypes.

Table A1: Balance Test: Students in the sample and Students in the province

	BG Province	Sample	Diff	Norm. Diff
Girl	0.497	0.510	-0.013	-0.026
Nursery School	0.198	0.099	0.099***	0.281
Pre-School	0.899	0.784	0.115***	0.321
Foreigner	0.254	0.329	-0.075***	-0.165
Graduated Mother	0.157	0.078	.079***	0.246
Graduated Father	0.100	0.0459	.054***	0.209
Math Test Score	0.004	-0.054	0.054	0.054
Reading Test Score	0.002	-0.022	0.027	0.024

Notes: data from the standardized test score INVALSI of all Italian students in grade II and V. Students in the sample are those for which we have information on the standardized test scores in mathematics and the IAT of their math teacher and/or the standardized test scores in reading and the IAT of their humanities teacher.

Table A2: Summary Statistics of Teachers by Grade and Field (T-test analysis)

Panel 1: Science Teachers					
	mII	mV	t stat	diff	p value
Graduated	0,304	0,292	0,093	0,013	0,926
Seniority >20	0,478	0,625	-1,000	-0,147	0,323
Pedagogical high school	0,870	0,917	-0,512	-0,047	0,612
Age	48,609	48,125	0,176	0,484	0,861
IAT score	0,369	0,228	1,006	0,142	0,320
Panel 2: Humanities Teachers					
	mII	mV	t stat	diff	p value
Graduated	0,464	0,360	0,760	0,104	0,451
Seniority >20	0,429	0,600	-1,241	-0,171	0,220
Pedagogical high school	0,786	0,840	-0,499	-0,054	0,620
Age	46,464	47,960	-0,559	-1,496	0,579
IAT score	0,383	0,415	-0,237	-0,033	0,813

Notes: The table shows descriptive statistics on teachers, separately for science and humanities teachers, for the following variables: share of graduated teachers, number of years of teachings, share of teachers who have attended pedagogical high school, number of years of teaching in the same class, age, IAT score. The sample includes 78 unique teachers.

Table A3: The impact of Teachers' Stereotypes, Model 1

VARIABLES	(1)	(2)	(3)	(4)
	Math Grade II	Reading Grade II	Math Grade V	Reading Grade V
Girl	-0.0403 (0.694)	-0.210 (0.412)	0.137 (0.621)	-1.08e-05 (0.762)
Teacher Stereotypes*Girls	-0.293*** (0.0826)	0.0978 (0.0675)	-0.0217 (0.0617)	-0.0329 (0.0798)
Foreigner	-0.533*** (0.112)	-0.694*** (0.0803)	-0.486*** (0.129)	-0.597*** (0.145)
Graduated Mother	0.310** (0.143)	0.445*** (0.0986)	-0.109 (0.137)	0.587*** (0.163)
Graduated Father	0.585*** (0.116)	0.00421 (0.306)	0.462*** (0.107)	0.169 (0.142)
Nursery School	0.160 (0.163)	-0.136 (0.113)	0.166 (0.137)	-0.102 (0.131)
Pre-School	0.0521 (0.203)	0.206 (0.128)	0.0659 (0.218)	0.348*** (0.0982)
Graduated Teacher*Girls	-0.0863 (0.192)	-0.0291 (0.115)	-0.0206 (0.236)	0.0320 (0.301)
Seniority>20*Girls	-0.228 (0.262)	-0.362** (0.135)	0.0947 (0.228)	0.138 (0.256)
Age Teacher*Girls	0.00420 (0.0136)	0.00776 (0.00815)	-0.00630 (0.0121)	0.00482 (0.00917)
Pedagogical high school T*Girls	-0.0191 (0.169)	0.0767 (0.114)	-0.173 (0.103)	-0.0697 (0.338)
Observations	515	622	523	600
R-squared	0.291	0.217	0.258	0.239
Class FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls*Girl	Yes	Yes	Yes	Yes

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 29 in column 1, and 25 in column 2, 28 in column 3, and 26 in column 4). Dependent variable: Test score in Math (columns 1-3) and Reading (columns 2-4), for students in Grade II and V. We include the interactions between teacher controls and the girl dummy. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Table A4: Robustness Check I

Panel 1: Excluding the lowest and highest percentile				
	(1)	(2)	(3)	(4)
	Math	Reading	Math	Reading
	Grade II	Grade II	Grade V	Grade V
Girl	-0.0455	-0.208	-0.0294	-0.767
	(0.695)	(0.413)	(0.650)	(0.638)
Teacher Stereotypes*Girls	-0.271***	0.0903	0.00828	-0.0232
	(0.0763)	(0.0623)	(0.0518)	(0.0730)
Observations	515	622	504	547
R-squared	0.291	0.217	0.270	0.233
Panel 2: using a dummy variable (IAT>0)				
	(1)	(2)	(3)	(4)
	Math	Reading	Math	Reading
	Grade II	Grade II	Grade V	Grade V
Girl	-0.0174	-0.276	0.170	0.00698
	(0.742)	(0.382)	(0.591)	(0.765)
IAT>0*Girls	-0.443***	0.180	0.0716	-0.0835
	(0.140)	(0.123)	(0.133)	(0.199)
Observations	515	622	523	600
R-squared	0.288	0.218	0.258	0.239
Class FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls*Girl	Yes	Yes	Yes	Yes

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. In panel 1 we exclude highest and lowest percentile of the IAT test score, while in panel 2 we use a dummy variable instead of a continuous variable for the IAT test score. (IAT>0) is a dummy variable equal to 1 if the standardized IAT test score is larger than 0. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Table A5: The role of explicit stereotypes

VARIABLES	(1)	(2)	(3)	(4)
	Math Grade II	Reading Grade II	Math Grade V	Reading Grade V
Girl	-0.349 (0.726)	-0.405 (0.388)	0.0139 (0.478)	0.136 (0.728)
Teacher Expl. Stereotypes *Girls	-0.429** (0.194)	0.123 (0.240)	-0.229 (0.144)	0.421** (0.172)
Observations	515	622	523	600
R-squared	0.290	0.217	0.260	0.244
Class FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls*Girl	Yes	Yes	Yes	Yes

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. The explicit stereotypes variable is a dummy variable equal to 1 for those teachers who report a high level of agreement with the statement "Being a housewife allows a woman to fulfil herself as much as having a job". Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Table A6: Robustness Check (Controlling for Explicit Stereotypes)

VARIABLES	(1)	(2)	(3)	(4)
	Math Grade II	Reading Grade II	Math Grade V	Reading Grade V
Girl	-0.0853 (0.602)	-0.302 (0.413)	-0.145 (0.522)	0.139 (0.724)
Teacher Stereotypes*Girls	-0.235*** (0.0714)	0.0931 (0.0660)	-0.0664 (0.0600)	0.00805 (0.0816)
Teacher Stereotypes Expl	0.577** (0.243)	0.719*** (0.196)	0.111 (0.267)	-0.492*** (0.0944)
Teacher Stereotypes Expl*Girls	-0.331* (0.183)	0.109 (0.231)	-0.283* (0.161)	0.428** (0.171)
Observations	515	622	523	600
R-squared	0.296	0.218	0.261	0.244
Class FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls*Girl	No	No	No	No

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Table A7: Robustness Check (Controlling for Stereotypes of other teachers)

VARIABLES	(1)	(2)	(3)	(4)
	Math Grade II	Reading Grade II	Math Grade V	Reading Grade V
Girl	0.908*	-0.633	0.535	0.245
	(0.483)	(0.657)	(0.653)	(1.420)
Teacher Stereotypes *Girls	-0.275***	0.180*	-0.0584	0.100
	(0.0848)	(0.104)	(0.0819)	(0.189)
Other Teacher Stereotypes	-2.325*	-0.546	-0.829***	-0.192
	(1.341)	(1.062)	(0.283)	(0.194)
Other Teacher Stereotypes *Girls	0.184	-0.103	0.0893	-0.120
	(0.109)	(0.0893)	(0.0779)	(0.151)
Observations	458	457	380	384
R-squared	0.299	0.220	0.258	0.234
Class FE	Yes	Yes	Yes	Yes
Students Controls	Yes	Yes	Yes	Yes
Teacher Controls*Girl	No	No	No	No

Notes: OLS estimates. Standard errors in parentheses clustered at the teacher level (the number of teachers is 23 in column 1, 28 in column 2, 24 in column 3, 25 in column 4). Dependent variable: Test score in Math/Reading. Other Teachers refers to Humanities Teachers in column 1 and 3, while it refers to math teachers in columns 2 and 4. Student controls include dummies for foreigners, for attending nursery school and pre-school, for graduated mothers and graduated fathers. Teacher controls include dummies for graduated teachers, for having a pedagogical high school diploma, for the age and the seniority of the teacher.

Table A8: Questions on Explicit Stereotypes

Being a housewife allows a woman to fulfil herself as much as having a paid job
On the whole, men make better political than women do
A university education is more important for a boy than a girl
On the whole, men make better business executives than women do
In conditions of job scarcity, men should have more right to obtain it than women
There are innate biological differences in the mathematical abilities between women and men
Boys are better at math, girls at reading
When a mother works for pay, the children suffer
A job is alright but what most women really want is a home and children
All in all, family life suffers when the woman has a full-time job
A man's job is to earn money; a woman's job is to look after the home and family

Notes: The table reports item-based questions used to evaluate the level of explicit stereotypes of teachers.