Rev Econ Household
https://doi.org/10.1007/s11150-026-09841-5

Check for updates

Gender stereotypes and children's performance in primary schools

Piera Bello[1] · Annalisa Cristini[1] · Costanza Marconi[1] · Federica Origo[1]

Received: 31 July 2025 / Accepted: 9 March 2026

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. 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.

1 Introduction

In recent decades, women's educational attainment and employment have both increased more rapidly than men's. 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 behind 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 & Sand, 2018; Guiso et al., 2008; Else-Quest et al., 2010).

✉ Piera Bello
piera.bello@unibg.it

[1] University of Bergamo, Department of Economics, Bergamo, Italy

Published online: 30 March 2026

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 & Mendolia, 2017). This partly explains why most research on this topic focuses on secondary education (Niederle & Vesterlund, 2010; Rodriguez-Planas & Nollenberger, 2018; Burgess et al., 2022; Ellison & Swanson, 2023).

However, 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 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) relies on implicit stereotypes measured through the IAT and Alan et al. (2018) elicit explicit beliefs through survey questions, our contribution is to extend the use of the IAT to teachers, combined with questions on explicit stereotypes, in primary schools.

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). Higher IAT

values indicate a stronger male-math association, while negative values indicate stronger female-math associations. 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 (that is, with a stronger male-math association) exhibit a 0.29 larger gender gap in math score. No significant effects are found in Grade V, even though the direction of the effect is negative as expected. Interestingly, the educational background of teachers, especially education in pedagogical methods acquired when they were in high school, 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 female-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 male-math association).

In addition, although parental education does not mediate the effect of the teacher's implicit stereotypes on student achievement, 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.

Our results are broadly consistent with prior work, but we also identify new patterns that enrich the existing evidence. Previous 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 (Carlana, 2019). Interestingly, this is due to girls performing worse when assigned to a stereotyped teacher (i.e., with a strong male-math association), 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 2018) 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 high-school teachers discriminate against male students in all subjects. Avitzour et al. (2021) 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. Moreover, our contribution to the literature stems in two important patterns that diverge from these papers. First, we find strong grade heterogeneity: the effects of teachers' implicit stereotypes are concentrated in grade 2, while they are absent in grade 5, suggesting that early primary school is a particularly sensitive period for the formation of gendered beliefs and achievement gaps. Second, we show that, in the early years of education, the widening of the gender gap in math is driven by gains for girls when matched with counter-stereotypical teachers—those who associate girls with math, rather than by losses for girls with stereotypical teachers. This result points to the importance of exposing girls to positive role models in early schooling environments, in contrast to significant harm produced by biased teachers in pre-adolescence.

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 of April.[1] 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.[2]

School hours may range from 24–27 to 30 h per week, reaching 40 h 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.[3] 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 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.

Teaching activities are organized in classes, i.e., a group of children that stay together during all the five years of primary school.[4] By law, the class size ranges from 15 to 26–27 students. Smaller classes are allowed in very small municipalities or in the 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.[5]

Regarding the allocation of teachers to classes, while not fully random, is largely driven by institutional rules and administrative constraints rather than by systematic or strategic matching between teachers and students. Within each school, class

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] We inspected such documents for the primary schools in our sample. They report similar criteria for class formation.

groups are typically formed *before* teachers are assigned, and assignment decisions predominantly follow seniority-based rules ("anzianità di servizio"), internal scheduling needs, and the aim of ensuring coverage of subject specializations. Moreover, Italian primary schools place strong emphasis on *pedagogical continuity*: once assigned to a class, teachers usually remain with the same cohort for multiple years, and reassignments occur mainly due to retirements, transfers, or temporary leaves. Teachers with greater seniority may have priority in choosing their school, usually based on proximity and commuting time from their place of residence; however, they cannot choose the specific class within the school.

As a result, principals, teachers, and families have limited scope within school to tailor assignments to the characteristics or expected performance of specific students.

Taken together, these institutional features imply that—after controlling for school fixed effects or class fixed effects, and rich sets of student and teacher observables—the residual variation in exposure to teachers with different levels of gender stereotypes within the same school can be considered plausibly as-good-as-random. This supports our empirical strategy, which exploits within-teacher or within-school variation to identify the effect of teacher stereotypes on student outcomes.

Teachers in primary school are generalists, meaning that they do not specialize in a specific subject. A specific high-school pedagogical diploma[6] was required to become a primary school teacher until 2010; since then, teachers have to graduate with a specific 5-year university degree in primary education. The allocation of teachers to schools is administered through a centralized system on the basis of seniority, which influences, as mentioned above, 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 h per week. This implies that classes with short hours usually have one teacher for all subjects (except for English and Catholic religion, which are taught by specific teachers), while classes with 30 or 40 h 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.[7] 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

[6] The study plan of the psycho-pedagogical diploma includes up to 5 h per week of psychology and pedagogy in the last three years of the course.

[7] Since 2018, students in grade V take also a standardized test to assess their English proficiency.

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 5 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. On that occasion we explained the main objectives of the study, providing basic definitions of both 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[8] 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 in Spring 2022, from April 30 to June 17. 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.[9] 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.[10]

[8] See Greenwald, McGhee, and Schwartz (1998) and Corno, Burns, & La Ferrara (2018) for more information on the functioning of the test.

[9] 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.

[10] We link the data only for math and reading teachers. Note that a single teacher may teach multiple classes or cover both humanities and science subjects. In total, we are able to link IAT test scores to student data for 100 teacher-subject pairs. A unique teacher denotes an individual instructor, irrespective of the number of classes taught.

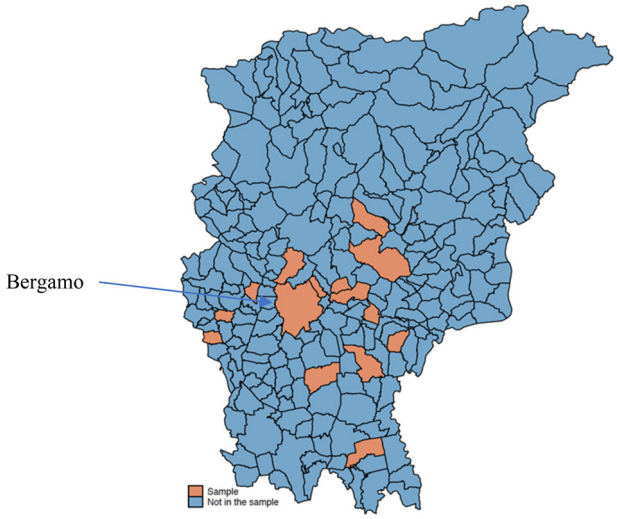


Fig. 1 Location of the schools included in our sample. Notes: the map reports the municipalities of the province of Bergamo. Municipalities highlighted in red are those included in our analysis

Figure 1 shows the location in the province of the schools in our final sample. The sampled schools are located in 16 different municipalities, including the province capital, mainly in the Center-South area of the province. Participating schools are on average larger than the average in the province (208 vs 162 pupils in the school year 2021–22).[11] Moreover, they are characterized by relatively lower socio-economic conditions, given the higher share of immigrants, the lower share of students with parents with a college degree and the lower share of students that attended nursery or pre-school (see Table 8 in the Appendix). However, the standardized difference for these variables is always below the threshold of 0.25 recommended by Imbens and Rubin (2015), with exception of the preschool and nursery variables. Interestingly, no statistically significant differences emerge for the share of girls and the standardized test scores. On the whole, these differences do not provide clear-cut evidence on school selection based on gender progressivity.

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-year-old and 50% of them have more than 20 years of teaching experience. Compared with the overall teaching staff in the province of Bergamo—where about 97% of primary school teachers are women—our sample shows a slightly smaller but still impactful female presence: 71 out of 78 teachers are women,

[11] This difference is largely explained by the lack of schools from the Northern part of the province, which is a mountains area with smaller and more sparse schools. Around 18.8% of the schools in our sample are located in mountain municipalities, compared to 29.6% in the province. If we exclude mountain municipalities, the average number of pupils is not statistically different between our sample and the provincial average (216 and 189 respectively).

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				

The table shows descriptive statistics (mean and standard deviation) 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. Total number of unique teachers: 78. Note that the same teacher can teach in more than one class

corresponding to roughly 91%[12]. This share is therefore broadly in line with the provincial average. In terms of age distribution, our sample includes a higher proportion of younger teachers than the province as a whole: 14% of respondents are under 34 (compared with 3.2% in the province). On the other hand, only 13% of sampled teachers are between 35 and 44, versus 23% in the Bergamo province. The share of older teachers in our sample is very similar to the provincial average: 41% of teachers are aged 45–54 (40% at the provincial level), while 32% are over 55 (compared with 30% provincially).

Regarding the IAT score, the estimates provided by psychological literature (Nosek & Smyth, 2011) suggest an average test score of 0.41; this implies that in our sample humanities teachers are relatively close to the average, as 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,

[12] Four of the remaining individuals in our sample are men and 3 chose not to disclose gender information. Furthermore, the low share of male teachers in primary school prevents from carrying out the analysis by teacher's gender.

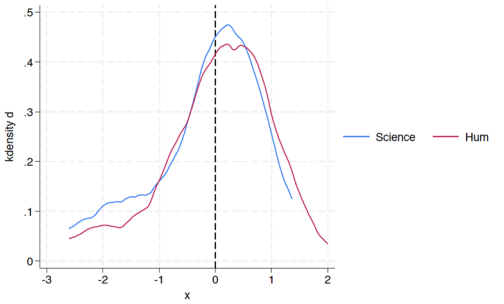


Fig. 2 Teachers' Implicit Gender Stereotypes by subject. Notes: This graph shows the distribution of Gender-Science IAT scores for science and humanities teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females

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 Fig. 6, 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.[13]

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 Fig. 2. 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 (see Table 9 in the appendix). No significant differences emerge.

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 scores 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.

[13] The figure shows that having a pedagogical high school diploma is associated with more than 1 s.d. 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	

The table shows descriptive statistics (mean and standard deviation) 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 city 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	571
Reading Test Score	−0.02	0.02	−0.53	−0.03	0.60	564	572
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

The table shows descriptive statistics and the difference between the genders in outcomes variables (the math test score and the reading test score), separately, for Grade II and Grade V

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 socio-economic status of students. It may be argued that parents can put pressure on school principals to assign their children to the 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 to the school principal.

Figure 7 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 foreign origin. The plot demonstrates that there is no systematic correlation between students' baseline characteristics and teacher stereotypes, with the exception of attending pre-school and being foreigner, which exhibit a weak correlation.

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:

$$y_{ic}\alpha_0 + \alpha_1(Femalei * IATc) + \alpha_2 Female_i + \alpha_3 X_i + \alpha_4(Female_i * Z_c) + \eta_c + \varepsilon_{ic} \tag{1}$$

where y_{ic} is the outcome (i.e., math or humanities standardized test score) of student i in class c. $Female_i$ 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), η_c are class fixed effects, 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 levels of stereotypes. This specification allows us 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:

$$\begin{aligned} y_{ics} = {} & \beta_0 + \beta_1(Female_i * IAT_c) + \beta_2 Female_i + \beta_3 IAT_c \\ & + \beta_4 X_i + \beta_5(Female_{i*} Z_c) + \beta_6 Z_c + \eta_s + \varepsilon_{ics} \end{aligned} \tag{2}$$

Where η_s are school fixed effects 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).

Table 4 The Impact of Teachers' Stereotypes, Model 1

VARIABLES	(1) Math	(2) Math	(3) Reading	(4) Reading
Girl	−0.149**	−0.352	0.159***	−0.0638
	(0.0609)	(0.561)	(0.0538)	(0.349)
Teacher Stereotypes*Girl	−0.0864	−0.104*	0.0340	0.0202
	(0.0605)	(0.0577)	(0.0698)	(0.0658)
Grade II	−0.0199	−0.0522	−0.267***	−0.288***
	(0.185)	(0.192)	(0.0886)	(0.0999)
Foreigner	−0.510***	−0.511***	−0.644***	−0.643***
	(0.0842)	(0.0846)	(0.0826)	(0.0824)
Graduated Mother	0.0707	0.0673	0.504***	0.499***
	(0.109)	(0.110)	(0.0877)	(0.0875)
Graduated Father	0.484***	0.481***	0.0931	0.101
	(0.0863)	(0.0902)	(0.162)	(0.159)
Nursery School	0.146	0.141	−0.128	−0.128
	(0.0958)	(0.0950)	(0.0855)	(0.0854)
Pre-School	0.0754	0.0714	0.281***	0.283***
	(0.140)	(0.141)	(0.0790)	(0.0790)
Graduated Teacher*Girl		0.00991		−0.0639
		(0.204)		(0.136)
Seniority>20*Girls		−0.251		−0.140
		(0.218)		(0.134)
Age Teacher*Girls		0.00794		0.00707
		(0.0108)		(0.00561)
Pedagogical high school T*Girls		−0.0507		−0.0203
		(0.143)		(0.167)
Observations	1038	1038	1222	1222
R-squared	0.261	0.263	0.221	0.222
Class FE	Yes	Yes	Yes	Yes
N Class	60	60	68	68

OLS estimates based on Equation 1. The variable Girl is a dummy equal to 1 if the student is female. Foreigner equals 1 if the student is foreign-born. Grade II is a dummy equal to 1 for students enrolled in the second grade. Graduated Mother and Graduated Father are dummies equal to 1 if the mother or the father, respectively, holds a university degree. Nursery School and Pre-School are indicators for having attended nursery school or pre-school prior to primary education. At the teacher level, Graduate Teacher is a dummy equal to 1 if the teacher holds a university degree; Seniority > 20 equals 1 if the teacher has more than 20 years of teaching experience; Age Teacher measures the teacher's age in years; Pedagogical High School T is a dummy equal to 1 if the teacher holds a pedagogical high school diploma. 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). The teacher stereotype variable is measured by the teacher's IAT test score. Column 1 and 2 refer to Math test, while columns 3-4 refer to Reading test. In all columns we include class FE

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 report the estimates of the model without the

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

VARIABLES	(1) Math Grade II	(2) Math Grade II	(3) Math Grade II	(4) Math Grade II
Girl	−0.0407	0.00775	−0.0403	−0.328
	(0.0807)	(0.0831)	(0.694)	(0.687)
Teacher Stereotypes*Girl	−0.300***	−0.299**	−0.293***	−0.263**
	(0.0921)	(0.106)	(0.0826)	(0.102)
Observations	515	515	515	515
R-squared	0.218	0.289	0.291	0.305
N Classes	28	28	28	28
VARIABLES	(1) Reading Grade II	(2) Reading Grade II	(3) Reading Grade II	(4) Reading Grade II
Girl	0.0391	0.0589	−0.210	−0.544
	(0.0756)	(0.0663)	(0.412)	(0.420)
Teacher Stereotypes*Girl	0.109	0.103	0.0978	0.135**
	(0.0870)	(0.0715)	(0.0675)	(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

OLS estimates based on Equation 1. 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. The variable Girl is a dummy equal to 1 if the student is female, while the Teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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

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 confirms the existence of gender gaps, with girls scoring −0.15 standard deviations lower than boys in math, while in reading, girls outperform boys with a score that is 0.16 standard deviations higher, keeping constant the other controls. 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,

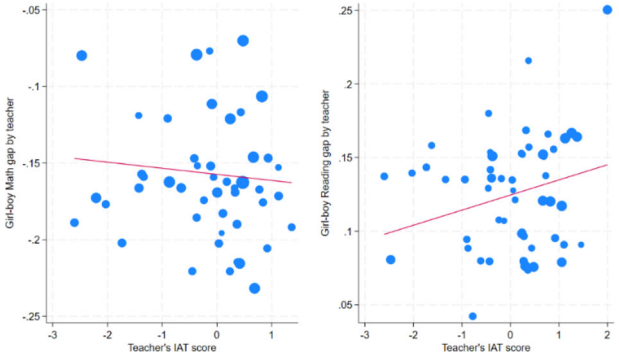


Fig. 3 Teachers' Implicit Gender Stereotypes and Gender Gap in Students' Performance (Math and Reading). Notes: The figure shows the correlation between teachers' IAT scores (where higher values indicate a stronger male–math association) and the gender gap in math and reading test scores at the teacher level. In panel (a), the negative relationship indicates that boys outperform girls in math when taught by more stereotyped teachers. In panel (b), the positive relationship indicates that girls outperform boys in reading when taught by more stereotyped teachers. The fitted line in the figure corresponds to a regression coefficient of −0.004 (standard error of 0.006) for panel (a) and 0.01* (standard error of 0.005) for panel (b), where * denotes statistical significance at the 10% level. Coefficients come from a second-stage regression of teacher-specific gender gaps on teachers' IAT scores: $y_s = \gamma_0 + \gamma_1 \, IAT_c$, where y_s denotes the teacher-level gender gap estimated in a first-stage hierarchical linear mixed model. In the first stage, we compute empirical Bayes estimates of teacher-specific gaps, which shrink estimates for teachers with relatively few students toward the overall mean gap. The coefficient γ_1 therefore captures the effect of teacher c's implicit stereotypes on the gender gap in performance among his/her students. Observations are weighted by the number of students per teacher

while fathers' one influencing their math performance.[14] 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.

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 male-math association of the teacher negatively affects the relative math performance of girls as compared to boys.

To corroborate the evidence from the student-level analysis, we replicate the exercise at the teacher–class level. Specifically, for each teacher in our sample, we compute the average difference between boys' and girls' test scores within the same class and relate this teacher-specific gender gap to the teacher's IAT score. Figure 3 plots the relationship between these variables. The girl–boy reading gap is positive and increases with the IAT score, while the girl–boy math gap is negative and, in absolute value, widens as the IAT score rises. The fitted line in the figure corresponds to a regression coefficient of −0.004 (standard error 0.006) for panel (a) and 0.01 (standard error 0.005) for panel (b). Thus, teachers with a stronger implicit association of math with males display larger gender gaps within their classes: girls perform relatively better than boys in reading, whereas boys perform relatively better

[14] A similar result is also found in Carlana & Corno (2024).

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

VARIABLES	(1) Math Grade V	(2) Math Grade V	(3) Math Grade V	(4) Math Grade V
Girl	−0.256***	−0.271***	0.137	0.0944
	(0.0841)	(0.0808)	(0.621)	(0.634)
Teacher Stereotypes*Girl	0.0260	−0.0273	−0.0217	−0.0252
	(0.0534)	(0.0516)	(0.0617)	(0.0625)
Observations	523	523	523	523
R-squared	0.208	0.257	0.258	0.261
N Classes	32	32	32	32
VARIABLES	(1) Reading Grade V	(2) Reading Grade V	(3) Reading Grade V	(4) Reading Grade V
Girl	0.189**	0.265***	−1.08e-05	−0.0755
	(0.0809)	(0.0790)	(0.762)	(0.814)
Teacher Stereotypes*Girl	0.0630	−0.0129	−0.0329	−0.0498
	(0.0904)	(0.0895)	(0.0798)	(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

OLS estimates based on Equation 1. 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. The variable Girl is a dummy equal to 1 if the student is female, while the Teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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

than girls in math. These results are coherent with the findings obtained from the student-level specification.

We return to the student-level framework to explore heterogeneity in the effects across grades.

In Tables 5 and 6, we estimate equation (1) by grade. 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.[15] In both tables, in Column 1, we provide the results of a

[15] Moreover, in Table 10 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.

Table 7 Is the Effect Due to Boys or Girls? Model 2

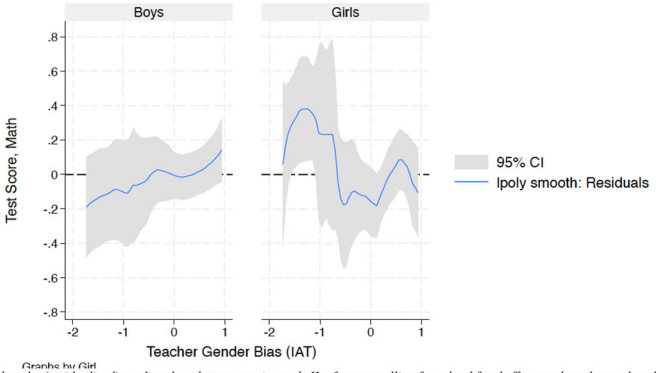
VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V	(5) Math	(6) Reading
Girl	0.00969	0.0446	−0.273***	0.259***	−0.0701	−0.0519
	(0.0814)	(0.0653)	(0.0801)	(0.0787)	(0.410)	(0.347)
Teacher Stereotypes	0.159	0.113	0.0443	0.0716	−0.0737	−0.0311
	(0.143)	(0.0710)	(0.0710)	(0.0564)	(0.0620)	(0.0520)
Teacher Stereotypes*Girl	−0.307***	0.106	−0.0479	−0.0188	−0.0645	−0.0321
	(0.105)	(0.0725)	(0.0493)	(0.0889)	(0.0527)	(0.0692)
Grade II					−0.161*	0.0295
					(0.0909)	(0.0747)
Teacher Stereotypes*Grade II					0.254***	0.164**
					(0.0830)	(0.0746)
Girl*Grade II					0.273**	−0.248**
					(0.109)	(0.103)
Teacher Stereotypes*Girls*Grade II					−0.256**	0.106
					(0.111)	(0.109)
Observations	515	622	523	600	1038	1222
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

OLS estimates based on Eq. 2. 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. The variable Girl is a dummy equal to 1 if the student is female, while the Teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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 age and seniority of the teacher

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 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 male-math stereotypes are positively associated with the relative performance of girls in reading. However, this effect is not precisely estimated in the other columns.

In Table 7, we present the estimates of Eq. 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



Notes: The figure plots the (residualized) predicted math test score in grade II, after controlling for school fixed effects and teacher and student characteristics, along the IAT score of the teacher (higher values indicate a stronger male–math association).

Fig. 4 The effect of teacher bias on student performance in math by gender, Grade II. Notes: The figure plots the (residualized) predicted math test score in grade II, after controlling for school fixed effects and teacher and student characteristics, along the IAT score of the teacher (higher values indicate a stronger male–math association)

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 that positively benefit boys' math performance or by the same teachers harming girls in this subject.

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 positive but not statistically significant. This suggests that male students are not harmed nor are they facilitated from having teachers with a stronger association between boys and math. However, the larger and negative coefficient of the interaction term suggests that the negative influence of stereotyped teachers on girls is pronounced.

In Fig. 4, we better explore this point and plot the predicted test score estimated using Eq. 2 over the distribution of the teachers' IAT test scores. According to the figure, no significant effect is found for boys along all the distributions 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 female-math association. Overall, these estimates highlight that the 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 unconventional teachers.

6 Further estimates

6.1 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 11 (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 male-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 and a IAT score greater than zero and the girl dummy. The results are provided in Table 11 (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).

Third, we check that our results are robust to alternative clustering levels and to specifications without clustering. Specifically, in Table 12 we report the results of Equation (1), for grade II, estimated using robust standard errors without clustering at the teacher level. In Table 13, we instead present the results of Eq. (2), which exploits variation in the performance of students of the same gender across classes assigned to teachers with different levels of stereotypes, using standard errors clustered at the school level. Although standard errors slightly increase when clustering at the school level, our main results remain statistically significant.

Fourth, we re-estimate the most complete specification of the model in Equation (1) leaving out one teacher at a time, to ensure that the effect we found on students' scores is not driven by some extreme values in IAT scores. Figure 8 shows the distribution of coefficients for both Math and Reading scores in Grade II and V: results confirm that the average coefficient in a leave-one-teacher-out specification mirrors the value of our main coefficient, suggesting that the core interaction does not depend on outliers.

Fifth, 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). [16] If we re-estimate Eq. 2 by subject and grade using this variable instead of the standardized

[16] For all the other statements on explicit stereotypes, less than 10% of the teacher's state that they agree with the statement. The full list of statements is in Table 17.

IAT, our baseline results are confirmed (see Table 14 in Appendix). While these estimates could imply that explicit stereotypes may be a good proxy for implicit ones, the correlation between explicit stereotypes and the IAT, though positive, is very low (0.05) and not statistically significant, as highlighed by the binscatter in Fig. 9. We then test if our main results survive when we control for explicit gender stereotypes. In particular, we re-estimate Eq. 2 including the dummy for explicit gender stereotypes and its interaction with the Girl dummy as additional regressors. The results are reported in Table 15. 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 15). 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.

Finally, an additional concern may be related to the fact that, when we do not control for class fixed effects (as in Eq. 2), the teacher's implicit stereotype may capture the overall teachers' attitudes rather than the behavior of the teacher of that specific subject. If this is true, effective interventions would require to measures and address gender stereotypes of all the teachers associated to a certain class rather than those of the teacher of the subject(s) characterized by large gender gaps in achievement (e.g, usually math). To test this hypothesis, we re-estimate Eq. 2 controlling for the standardized IAT of the other teacher (that is, the Humanities teacher in the math equation and the Science teacher in the reading equation). Estimates reported in Table 16 highlight that the gender gap in math in grade II is driven solely by the math teacher (the interaction term between student's gender and the IAT of the other teacher is never statistically significant).

6.2 Heterogeneous effects: the role of teacher's education background

Since descriptive evidence shown in Table 9 suggests that teachers' education background might be a determinant of gender stereotypes, we investigate whether it also influences how gender stereotypes affect the gender gap in students' performance. Specifically, we expect that teachers with a strong pedagogical background may be better equipped to adapt their teaching methods to the needs of different students, using a student-centered approach rather than a subject-centered approach.[17] This should contribute to reducing 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 because, as discussed in Section 2, 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 Eq. 2. Results are reported in Tables 18 and 19, respectively. Interestingly, Table 18 shows that the attendance of a

[17] 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).

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 Stereoptypes∗Girls" is positive and statistically significant at 10 percent level in column 1). On the contrary, we find that having a graduated or non-graduated teacher has no effect on the math gender gap in Grade II (the coefficient of the variable labelled "Graduated T∗Teacher Stereotypes∗Girls" is not statistically significant in column 1 of Table 19).[18] 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).

Overall, these findings highlight the importance of prospective 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 teacher's 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 heterogenous effects by other teacher and student characteristics. Specifically, we analyzed the effect of years of exposure to stereotyped teachers (Table 20) 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 mediating 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.[19]

Although the heterogeneity analysis provides interesting evidence, we acknowledge its limitations due to the small sample size. The number of teachers in our data is relatively small, which restricts the statistical power of subgroup comparisons. As a consequence, the heterogeneity patterns we document should be viewed as exploratory rather than definitive. These estimates offer suggestive evidence on how the effects may vary across teacher characteristics, but they should be interpreted with caution given the limited sample size.

[18] 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.

[19] Estimates are available upon request.

7 Discussion and conclusion

7.1 Discussion

In the growing literature on the effects of teachers' gender stereotypes on student achievement, our study provides new evidence from an earlier and critical stage of schooling. While Carlana (2019) examines implicit stereotypes among middle-school teachers in Italy, we focus on primary school, a formative period in which students' beliefs, attitudes, and self-confidence begin to take shape and in which teachers play a key role in reinforcing or challenging gender norms that may originate in the family environment. To date, the closest study in this setting is Alan et al. (2018), who investigate the effects of teachers' explicit gender-role attitudes in Turkish primary schools. Our work differs in two important respects. First, rather than relying on self-reported attitudes, we measure teachers' gender beliefs using the Implicit Association Test (IAT), which is less susceptible to social desirability concerns—particularly salient for teachers, whose professional norms may make them reluctant to express stereotypical views explicitly even when such views are held implicitly. Second, our findings reveal two novel patterns. We document substantial grade heterogeneity: effects are strong in grade 2 but absent in grade 5, underscoring the heightened sensitivity of early primary school for the development of gendered beliefs and achievement gaps. Moreover, the widening of the gender gap in math does not arise from losses for girls taught by stereotypical teachers. Instead, it is entirely driven by gains for girls assigned to counter-stereotypical teachers—those who associate girls with math—highlighting the importance of positive role models in early schooling environments.

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 10, 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 in future research.

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 & Votruba-Drzal, 2011) and that it is mainly girls' math performance to be affected by the teacher-child relationship (McCormick & O'Connor, 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. 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.

Given the small dimension of our sample, generalization of our results might seem far-reaching. Also, the focus on one Italian province in one specific cohort may raise concerns on how comparable our setting is to the whole country.

To support the external validity of our results, we compare the characteristics of our sample with those of the average primary school in the province of Bergamo, as well as with regional and national benchmarks. Among the 1479 pupils included in our study, approximately 49% are female—exactly in line with the gender composition of primary public schools in the province (49%) and very close to the regional and national figures (48% in both Lombardy and Italy). Regarding students' origin, our sample is aligned with the provincial population: about 67% of pupils were born in Italy, compared with 78% at the provincial level, whereas the native over foreign ratio is slightly higher at provincial level (with 81% pupils in Lombardy schools being of Italian origin), and at national level (87% are native Italians). As discussed in Section 3, 91% of participating teachers in our experiment are women. This proportion mirrors the strong female predominance in primary education observed at higher aggregation levels: 97% in the province of Bergamo, and 96% both regionally and nationally.

On the whole, our sample does not differ significantly from the population of students and teachers, especially at the regional level, reassuring about external validity. Differences in student composition compared with the national average are not due to sample selection, but to the structural North-South divide (Ballarino et al., 2014).

7.2 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 the gender gap in student performance is significantly influenced by the implicit gender biases held by teachers, especially in early grades in primary school. In these years, the beneficial effects of matching girls with teachers characterized by "unconventional" gender stereotypes may be particularly salient: when girls are taught by teachers with a strong female-math association, their math performance tends to be better, a result in line with models documenting the role of stereotypes on the formation of beliefs and psychological attitude (Jouini et al., 2018). By analyzing the role of heterogeneity, we find that a pedagogical education

background is a potential channel to mitigate the impact of their stereotypes on students' performance. This is in line with the evidence from a small-scale Randomized Control Trial carried out in primary schools in Italy showing 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).

On the whole, our 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 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. It also calls for targeted interventions to improve educational outcomes and close the gender gap in math, which encompass raising teachers' awareness on their inherent biases and designing training courses to guarantee equitable treatment of all students.

Data Availability The data supporting the findings of this study were obtained by matching student-level INVALSI test scores with survey data collected from teachers. This matching was conducted using anonymized data and was authorized by school principals specifically for the purposes of this study. As a result, the data cannot be made publicly available.

Acknowledgements We are grateful to INVALSI, especially Patrizia Falzetti and Patrizia Giannantoni, and the provincial working group for economic recovery and women's empowerment for their support to data collection. We also thank the participants at the SABE-IAREP conference in 2023, the 2023 and 2025 INVALSI conferences, the EC-JRC seminar (February 2024), and the 2024 AIEL conference for their helpful comments.

Author contributions F.O. and A.C. contributed to the conceptualization of the research question and data curation. P.B. and C.M. carried out the formal analysis. All authors contributed to the writing of the paper.

Funding 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. Open access funding provided by Università degli Studi di Bergamo within the CRUI-CARE Agreement.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

8 Appendix

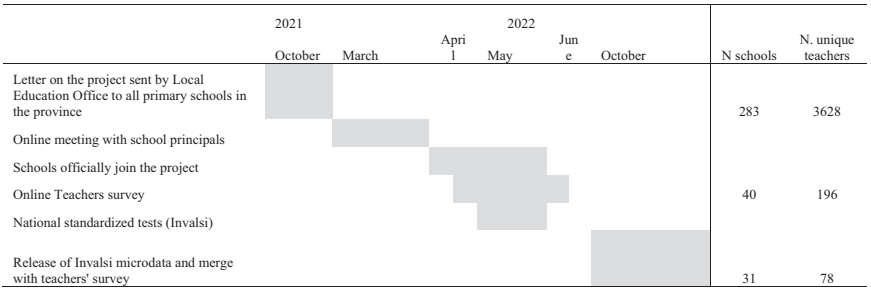


Fig. 5 Time line Project. Notes: The figure reports the timeline of the project from October 2021 to October 2022. The rightmost columns indicate the number of schools involved at each stage and the number of unique teachers observed in the corresponding phase. "N. unique teachers" refers to the number of distinct teachers (net of duplicates) participating in each activity

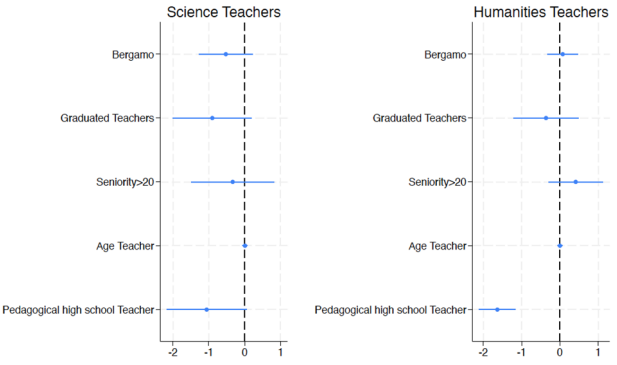


Fig. 6 Correlation between Gender IAT Score and Teachers characteristics by Field, Notes: The figure reports OLS estimates and 95% confidence intervals from regressions estimated separately for Science teachers (left panel) and Humanities teachers (right panel) and show the correlation between teachers' IAT score (higher values indicate a stronger male–math association) and own characteristics. The estimates come from a model with school FE and standard errors clustered at the school level. The vertical dashed line indicates zero. Bergamo is a dummy equal to 1 if the school is located in the province of Bergamo. Graduated Teachers is a dummy equal to 1 if the teacher holds a university degree. Seniority > 20 is a dummy equal to 1 if the teacher has more than 20 years of teaching experience. Pedagogical high school Teacher is a dummy equal to 1 if the teacher holds a pedagogical high school diploma. Age Teacher measures the teacher's age in years

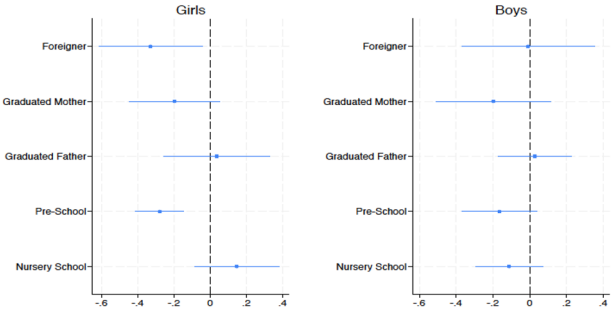


Fig. 7 Correlation between Gender IAT Score and students' characteristics. The figure reports OLS estimates and 95% confidence intervals from regressions estimated separately for girls (left panel) and boys (right panel). The coefficients capture the association between student characteristics and the outcome variable. The estimates come from a model including school fixed effects, and standard errors are clustered at the school level. The vertical dashed line indicates zero. Foreigner is a dummy equal to 1 if the student is foreign-born. Graduated Mother and Graduated Father are dummies equal to 1 if the mother or the father, respectively, holds a university degree. Pre-School and Nursery School are dummies equal to 1 if the student attended pre-school or nursery school prior to primary education

Figure of Teacher- – x Girl

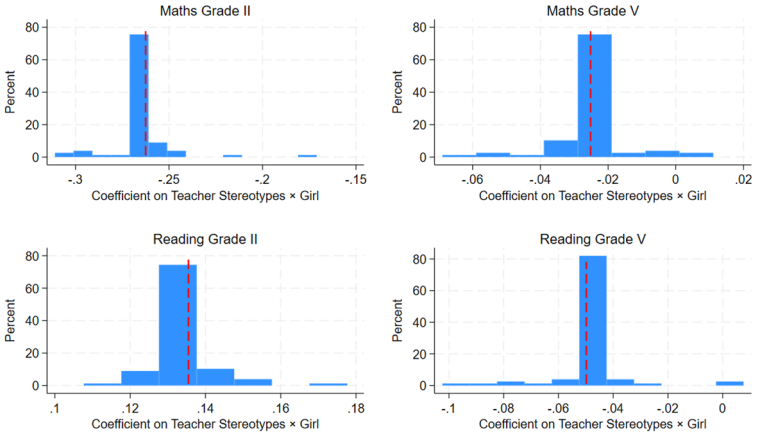


Fig. 8 Distribution of Leave-One-Teacher-Out Regression Coefficients – Teacher Stereotypes x Girl. Notes: The histograms show the distribution of coefficients on "Teacher Stereotypes x Girl" of Equation (1) including class fixed effects, and students, students x girls and teachers x girls controls, following a leave-one-teacher-out approach. The red dotted line indicates the value of the "Teacher Stereotypes x Girl" coefficient in the main regression

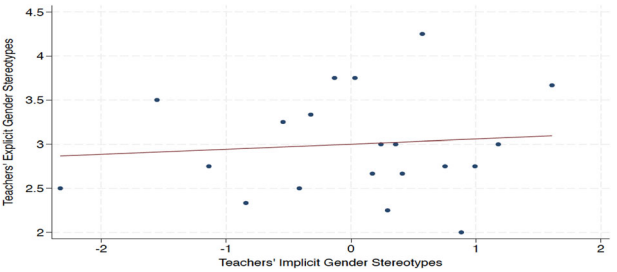


Fig. 9 Correlation between Explicit and Implicit Stereotypes. Notes: The binscatter shows the correlation between the explicit stereotype variable and the IAT score. For the teacher' IAT score higher values indicate a stronger male–math association. In order to measure the level of teacher's explicit 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 8 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	0.0.99	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

The table shows 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. *** indicates significance at the 1% level ($p < 0.01$)

Table 9 Summary Statistics of Teachers by Grade and Field (T-test analysis)

Panel 1: Science Teachers					
	mII	mV	t_stat	diff	p_value
Graduated	0304	0292	0093	0013	0926
Seniority >20	0478	0625	−1000	−0147	0323
Pedagogical high school	0870	0917	−0512	−0047	0612
Age	48,609	48,125	0176	0484	0861
IAT score	0369	0228	1006	0142	0320
Panel 2: Humanities Teachers					
	mII	mV	t_stat	diff	p_value
Graduated	0464	0360	0760	0104	0451
Seniority >20	0429	0600	−1241	−0171	0220
Pedagogical high school	0786	0840	−0499	−0054	0620
Age	46,464	47,960	−0559	−1496	0579
IAT score	0383	0415	−0237	−0033	0813

The table shows descriptive statistics and differences by grade 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 10 The Impact of Teachers' Stereotypes, Model 1

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
Girl	−0.0403	−0.210	0.137	−1.08e-05
	(0.694)	(0.412)	(0.621)	(0.762)
Teacher Stereotypes*Girl	−0.293***	0.0978	−0.0217	−0.0329
	(0.0826)	(0.0675)	(0.0617)	(0.0798)
Foreigner	−0.533***	−0.694***	−0.486***	−0.597***
	(0.112)	(0.0803)	(0.129)	(0.145)
Graduated Mother	0.310**	0.445***	−0.109	0.587***
	(0.143)	(0.0986)	(0.137)	(0.163)
Graduated Father	0.585***	0.00421	0.462***	0.169
	(0.116)	(0.306)	(0.107)	(0.142)
Nursery School	0.160	−0.136	0.166	−0.102
	(0.163)	(0.113)	(0.137)	(0.131)
Pre-School	0.0521	0.206	0.0659	0.348***
	(0.203)	(0.128)	(0.218)	(0.0982)
Graduated Teacher*Girls	−0.0863	−0.0291	−0.0206	0.0320
	(0.192)	(0.115)	(0.236)	(0.301)

Table 10 continued

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
Seniority>20*Girls	−0.228	−0.362**	0.0947	0.138
	(0.262)	(0.135)	(0.228)	(0.256)
Age Teacher*Girls	0.00420	0.00776	−0.00630	0.00482
	(0.0136)	(0.00815)	(0.0121)	(0.00917)
Pedagogical high school T*Girl	−0.0191	0.0767	−0.173	−0.0697
	(0.169)	(0.114)	(0.103)	(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

OLS estimates based on Equation 1. The variable Girl is a dummy equal to 1 if the student is female. Foreigner equals 1 if the student is foreign-born. Grade II is a dummy equal to 1 for students enrolled in the second grade. Graduated Mother and Graduated Father are dummies equal to 1 if the mother or the father, respectively, holds a university degree. Nursery School and Pre-School are indicators for having attended nursery school or pre-school prior to primary education. At the teacher level, Graduate Teacher is a dummy equal to 1 if the teacher holds a university degree; Seniority > 20 equals 1 if the teacher has more than 20 years of teaching experience; Age Teacher measures the teacher's age in years; Pedagogical High School T is a dummy equal to 1 if the teacher holds a pedagogical high school diploma. 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 11 Robustness Check I

Panel 1: Excluding the lowest and highest percentile

	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
Girl	−0.0455	−0.208	−0.0294	−0.767
	(0.695)	(0.413)	(0.650)	(0.638)
Teacher Stereotypes*Girl	−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) Math	(2) Reading	(3) Math	(4) Reading

Table 11 continued

Panel 2: Using a dummy variable (IAT > 0)				
	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
	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

OLS estimates based on Equation 1. 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). The variable Girl is a dummy equal to 1 if the student is female, while the Teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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 12 The Impact of Teachers' Stereotypes, Model 1, Grade II Without clustering

VARIABLES	(1) Math Grade II	(2) Math Grade II	(3) Math Grade II	(4) Math Grade II
Girl	−0.0407	0.00775	−0.0403	−0.328
	(0.0817)	(0.0796)	(0.749)	(0.786)
Teacher Stereotypes*Girl	−0.300***	−0.299***	−0.293**	−0.263**
	(0.114)	(0.110)	(0.116)	(0.129)
Observations	515	515	515	515
R-squared	0.218	0.289	0.291	0.305
N Classes	28	28	28	28
VARIABLES	(1) Reading Grade II	(2) Reading Grade II	(3) Reading Grade II	(4) Reading Grade II
Girl	0.0391	0.0589	−0.210	−0.544
	(0.0796)	(0.0763)	(0.552)	(0.605)

Table 12 continued

VARIABLES	(1) Reading Grade II	(2) Reading Grade II	(3) Reading Grade II	(4) Reading Grade II
Teacher Stereotypes*Girl	0.109	0.103	0.0978	0.135
	(0.105)	(0.0983)	(0.109)	(0.111)
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

OLS estimates based on Equation 1. Robust standard errors in parentheses. The variable Girl is a dummy equal to 1 if the student is female, while the Teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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

Table 13 Is the Effect Due to Boys or Girls? Model 2 Clustering at the school level

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V	(5) Math	(6) Reading
Girl	0.00969	0.0446	−0.273***	0.259***	−0.0701	−0.0519
	(0.0646)	(0.0677)	(0.0671)	(0.0686)	(0.423)	(0.250)
Teacher Stereotypes	0.159	0.113	0.0443	0.0716	−0.0737	−0.0311
	(0.214)	(0.0875)	(0.0954)	(0.0554)	(0.0815)	(0.0612)
Teacher Stereotypes*Girl	−0.307**	0.106	−0.0479	−0.0188	−0.0645	−0.0321
	(0.115)	(0.0769)	(0.0407)	(0.0825)	(0.0404)	(0.0589)
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

OLS estimates based on Eq. 2. Standard errors in parentheses clustered at the school level (the number of schools is 12 in column 1, 14 in column 2, 13 in column 3, 13 in column 4, 15 in column 5, and 16 in column 6). The variable Girl is a dummy equal to 1 if the student is female, while The teacher stereotypes variable refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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 age and seniority of the teacher

Table 14 The Role of Explicit Stereotypes

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
Girl	−0.349	−0.405	0.0139	0.136
	(0.726)	(0.388)	(0.478)	(0.728)
Teacher Expl. Stereotypes *Girl	−0.429**	0.123	−0.229	0.421**
	(0.194)	(0.240)	(0.144)	(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

OLS estimates based on Equation 1. 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). The variable Girl is a dummy equal to 1 if the student is female. 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". 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 15 Robustness Check (Controlling for Explicit Stereotypes)

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) Reading Grade V
Girl	−0.0853	−0.302	−0.145	0.139
	(0.602)	(0.413)	(0.522)	(0.724)
Teacher Stereotypes*Girl	−0.235***	0.0931	−0.0664	0.00805
	(0.0714)	(0.0660)	(0.0600)	(0.0816)
Teacher Stereotypes Expl	0.577**	0.719***	0.111	−0.492***
	(0.243)	(0.196)	(0.267)	(0.0944)
Teacher Stereotypes Expl*Girls	−0.331*	0.109	−0.283*	0.428**
	(0.183)	(0.231)	(0.161)	(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

OLS estimates based on Eq. 2. The variable Girl is a dummy equal to 1 if the student is female, while Teacher stereotypes refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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". 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 16 Robustness Check (Controlling for Stereotypes of other teachers)

VARIABLES	(1) Math Grade II	(2) Reading Grade II	(3) Math Grade V	(4) 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

OLS estimates based on Equation 1. The variable Girl is a dummy equal to 1 if the student is female, while Teacher stereotypes refers to the teacher' IAT score (where higher values indicate a stronger male–math association). 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". 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 17 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

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

Table 18 The Impact of the Teacher's Background (pedagogical high school) (Model 2)

VARIABLES	(1) Math II	(2) Reading II	(3) Math V	(4) Reading V
Girl	0.0624	0.0996	−0.114**	−0.477**
	(0.0531)	(0.0719)	(0.0490)	(0.213)
Teacher Stereotypes	0.418	−0.323***	0.299	−0.122*
	(0.575)	(0.0632)	(0.467)	(0.0706)
Teacher Stereotypes*Girl	−0.516***	0.112	0.127	0.729***
	(0.0771)	(0.0723)	(0.174)	(0.0963)
Pedagogical high school (HS) Teacher	−0.372*	0.0499	0.260	0.193*
	(0.206)	(0.0719)	(0.238)	(0.111)
Pedagogical HS*Teacher Stereotypes	−0.267	0.520***	−0.252	0.229***
	(0.707)	(0.0723)	(0.476)	(0.0681)
Pedagogical HS*Girl	−0.0628	−0.0674	−0.179*	0.727***
	(0.107)	(0.107)	(0.0972)	(0.226)
Pedagogical HS*Teacher Stereotypes*Girl	0.222*	−0.0170	−0.181	−0.826***
	(0.115)	(0.120)	(0.200)	(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

OLS estimates based on Eq. 2. The variable Girl is a dummy equal to 1 if the student is female, while Teacher stereotypes refers to the teacher' IAT score (where higher values indicate a stronger male–math association). Pedagogical High School T is a dummy equal to 1 if the teacher holds a pedagogical high school diploma. 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 19 The Impact of the Teacher's Education Level (graduated vs non graduated) (Model 2)

VARIABLES	(1) Math II	(2) Reading II	(3) Math V	(4) Reading V
Girl	0.00530	−0.00320	−0.309**	0.325***
	(0.0945)	(0.0888)	(0.111)	(0.0991)
Teacher Stereotypes	0.337***	0.183**	0.0889	0.0554
	(0.104)	(0.0736)	(0.0860)	(0.0612)
Teacher Stereotypes*Girls	−0.265***	0.136	−0.0950	0.0311
	(0.0930)	(0.109)	(0.0577)	(0.0924)
Graduated Teachers	−0.398***	−0.174	−0.221	0.294**
	(0.0718)	(0.116)	(0.242)	(0.133)
Graduated T*Teacher Stereotypes	−0.546***	−0.200	−0.209*	−0.00308
	(0.143)	(0.156)	(0.120)	(0.0666)
Graduated T*Girls	−0.0605	0.103	0.117	−0.133
	(0.132)	(0.125)	(0.124)	(0.157)
Graduated T*Teacher Stereotypes*Girl	−0.214	−0.0551	0.224**	−0.266
	(0.217)	(0.141)	(0.0992)	(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

OLS estimates based on Eq. 2. The variable Girl is a dummy equal to 1 if the student is female, while Teacher stereotypes refers to the teacher' IAT score (where higher values indicate a stronger male–math association). Graduate Teacher is a dummy equal to 1 if the teacher holds a university degree. 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 20 The Impact of the Length of Exposure (Exposure same teacher>2) (Model 2)

VARIABLES	(3) Math V	(4) Reading V
Girl	−0.152***	0.198*
	(0.0400)	(0.107)
Teacher Stereotypes	0.640	0.0810
	(0.580)	(0.113)
Teacher Stereotypes*Girls	0.0796	−0.0573
	(0.0502)	(0.128)
More than 2 years same class	0.158	0.0491
	(0.115)	(0.112)
More than 2 years same class*Teacher Stereotypes	−0.550	0.00419
	(0.535)	(0.120)
More than 2 years same class*Girls	−0.159	0.0765
	(0.107)	(0.146)
More than 2 years same*Teacher Stereotypes*Girl	−0.155*	0.0386
	(0.0809)	(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

OLS estimates based on Eq. 2. The variable Girl is a dummy equal to 1 if the student is female, while Teacher stereotypes refers to the teacher' IAT score (where higher values indicate a stronger male–math association). More than 2 years with the same teacher is a dummy variable equal to 1 if the student has been taught by the same teacher for more than two consecutive years, and 0 otherwise. 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

References

Alan, S., Ertac, S., & Mumcu, I. (2018) Gender Stereotypes in the Classroom and Effects on Achievement. *Review of Economics and Statistics*, *v. 100*(iss. 5), 876–90.

Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2021) "On The Origins Of Gender-Biased Behavior: The Role Of Explicit And Implicit Stereotypes." NBER Working Paper 27818.

Ballarino, G., Panichella, N., & Triventi, M. (2014) School expansion and uneven modernization. Comparing educational inequality in Northern and Southern Italy. *Research in Social Stratification and Mobility*, *36*, 69–86.

Barbieri, G., Rossetti, C., & Sestito, P. (2011) The determinants of teacher mobility: Evidence using Italian teachers' transfer applications. *Economics of education review*, *30*, 1430–1444.

Brindusa, A., Rodriguez-Planas, N., & Sanz-de-Galdeano, A. (2020) Is the math gender gap associated with gender equality? Only in low-income countries. *Economics of Education Review*, *v. 79*, 102064. December 2020.

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*, 102267.

Carlana, M. (2019) "Implicit Stereotypes: Evidence from Teachers' Gender Bias." The Quarterly Journal of Economics, 1163-224.

Carlana, M., & Corno, L. (2024) Thinking about Parents: Gender and field of study. *AEA Papers and Proceedings*, *114*, 254–8.

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.

Corno, L., Burns, J., & La Ferrara, E. "Interaction, Stereotypes and Performance. Evidence from South Africa." BREAD Working Paper no. 549, 2018.

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.

Dossi, G., Figlio, D., Giuliano, P., & Sapienza, P. (2021) Born in the family: Preferences for boys and the gender gap in math. *Journal of Economic Behavior and Organization*, *v. 183*, 175–88.

Ellison, G., & Swanson, A. (2023) Dynamics of the gender gap in high math achievement. *Journal of Human Resources*, *v. 58*(iss. 5), 1679–711.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010) Cross national patterns of gender differences in mathematics and gender equity. A meta analysis. *Psychol. Bull*, *136*, 103–127.

Ertl, B., Luttenberger, S., & Paechter, M. (2017) The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology*, *8*, 703.

Guiso, L., Monte, F., Sapienza, P., & Zingales, P. (2008) Culture, gender, and math. *Science*, *320*, 1164–5.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol*, *74*, 1464–80.

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*, 915–45.

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*, 34–45.

Lavy, V. (2008). Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment. Journal of Public Economics. DOI link.

Lavy, V., & Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *J. Public Econ*, *167*, 263–279.

Imbens, G. W., & Rubin, D. B. (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge university press

Maldonado-Carreño, C., & 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), 601–16.

McCormick, M. P., & O'Connor, E. E. (2015) Teacher–child relationship quality and academic achievement in elementary school: Does gender matter?. *Journal of Educational Psychology*, *107*(2), 502.

Nicoletti, C., Sevilla, A., Tonei, V. "Gender Stereotypes in the Family." (2022).

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), 129–44.

Nosek, B. A., & Smyth, F. L. (2011) Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125–56.

OECD (2009) Creating Effective Teaching and Learning Environments First Results from TALIS. Paris: OECD Publishing.

OECD (2014) Talis 2013 Results: An International Perspective on Teaching and Learning. Paris: OECD Publishing.

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*, 230–53.

Wigfield, A., et al. (1997) 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), 451.