Online Appendix

Implicit Stereotypes: Evidence from Teachers' Gender Bias

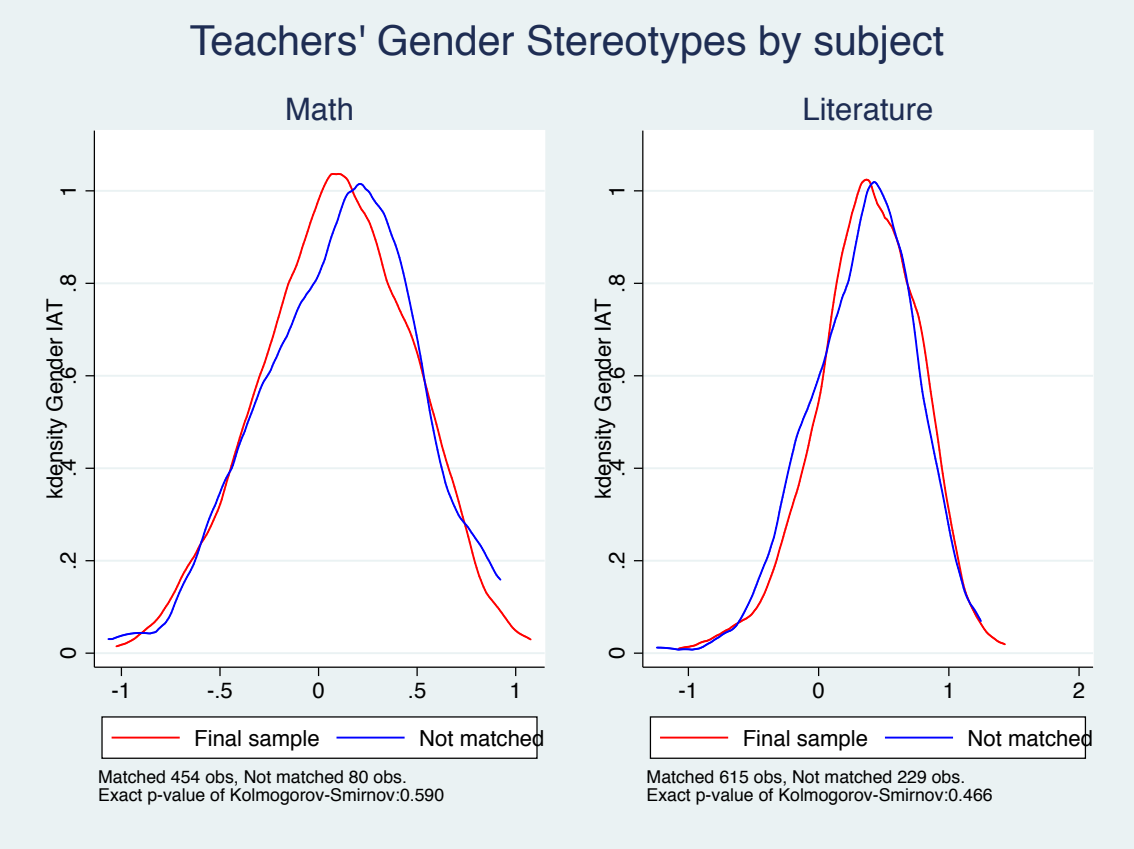
Michela Carlana*

January 2019

*Harvard Kennedy School (e-mail: michela_carlana@hks.harvard.edu).

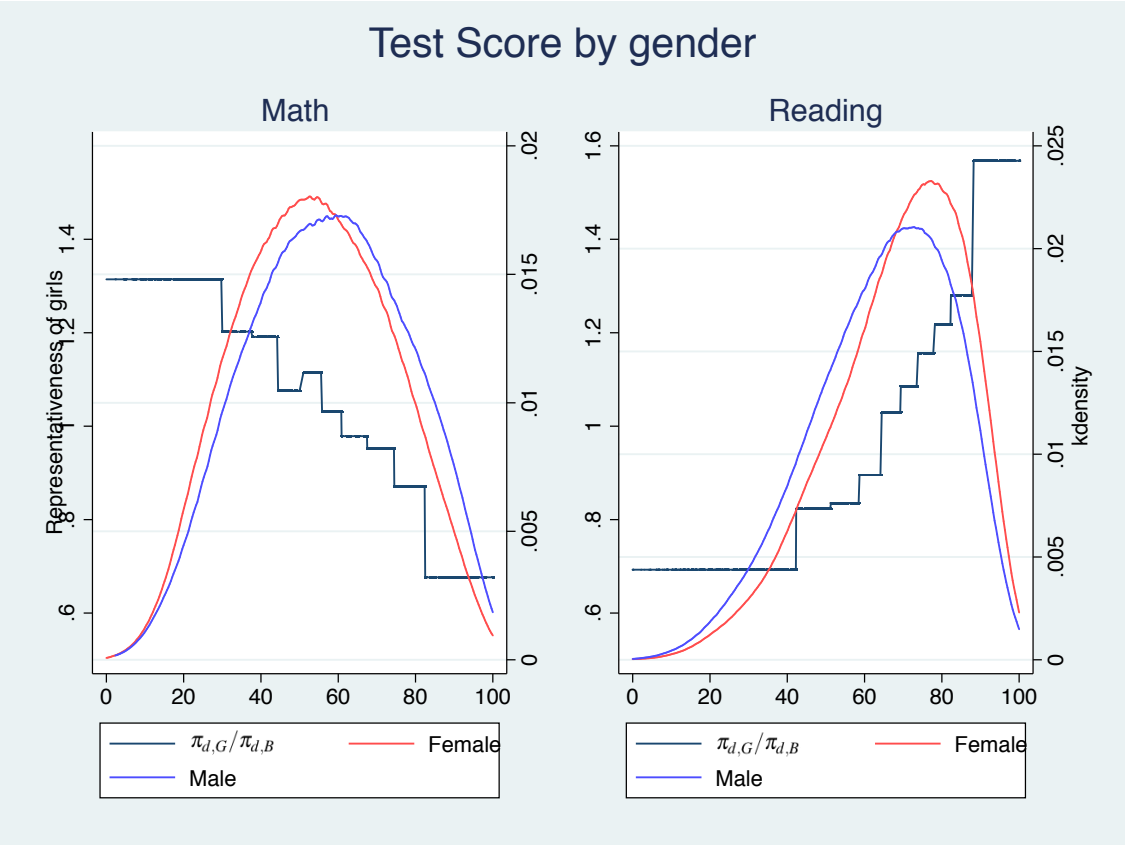
A Additional Figures and Tables

Figure A.I. Teachers' Implicit Gender Stereotypes (IAT measure) by subject of matched and unmatched sample



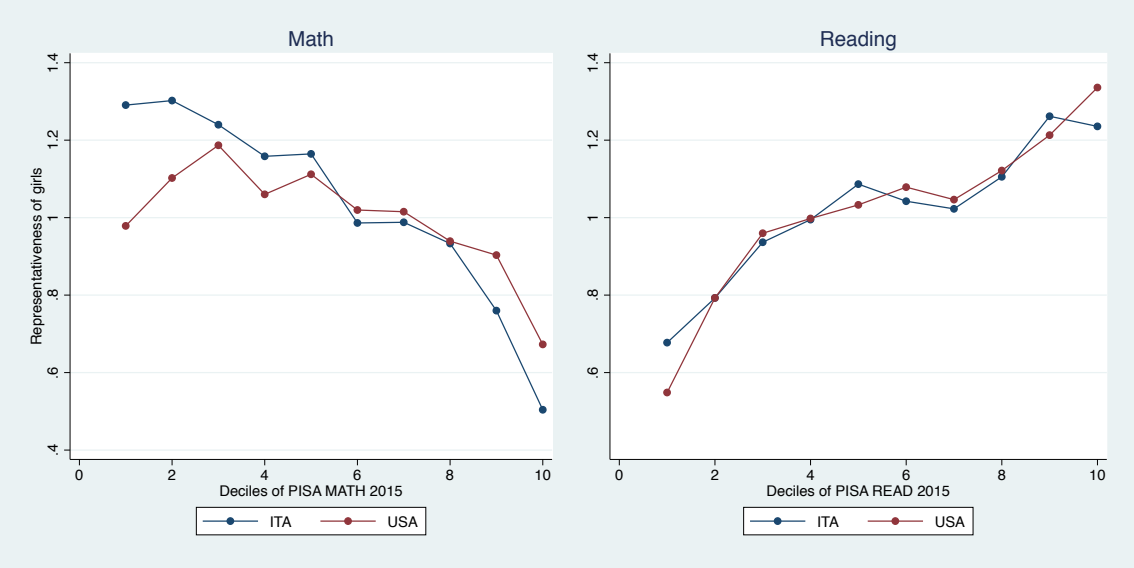
Notes: This graph shows the distribution of Gender-Science IAT scores for math and literature teachers. A higher value of implicit bias indicates a stronger association between scientific-males and humanistic-females. Zero indicates no gender stereotypes. The graph provides evidence that teachers used in this paper are similar in terms of gender stereotypes to those who completed the survey but are not included in the main analysis because I do not have outcomes of their students.

Figure A.II. Gender differences in math and reading standardized test score



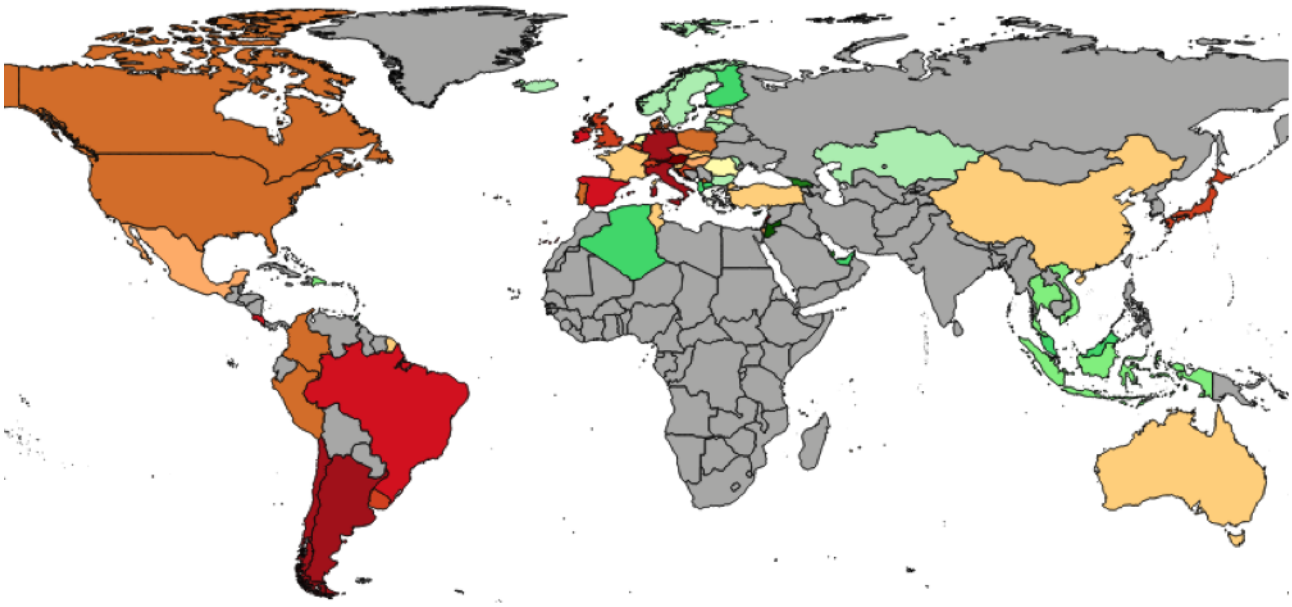
Notes: This graph shows the difference in math and reading test scores in grade 8 between females and males in the sample of Italian middle schools.

Figure A.III. Representativeness of girls with respect to boys in PISA 2015 in US and Italy



Notes: This graph shows the representativeness of girls with respect to boys in PISA 2015 (math and reading).

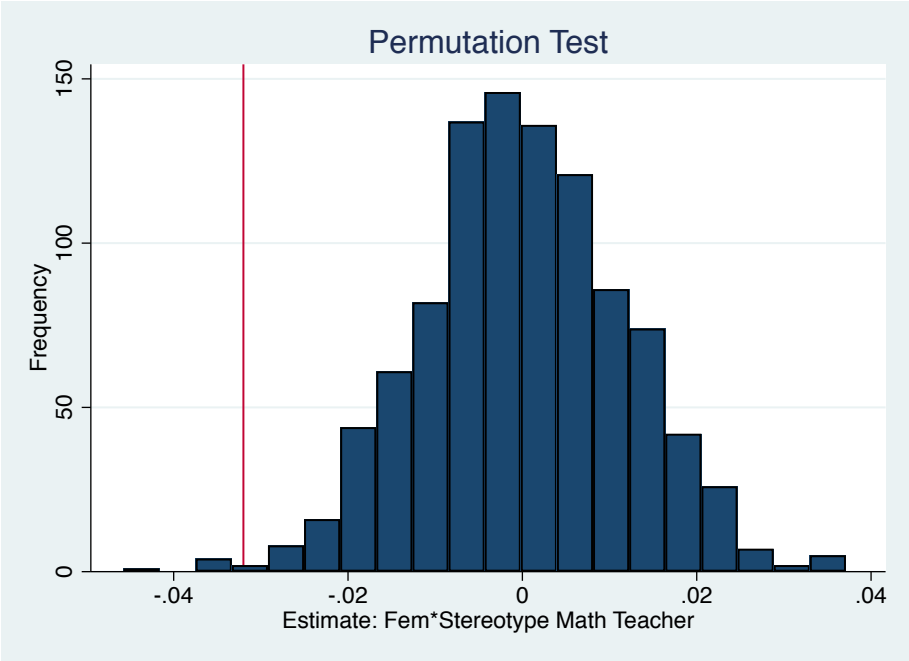
Figure A.IV. Gender differences in math standardized test score PISA



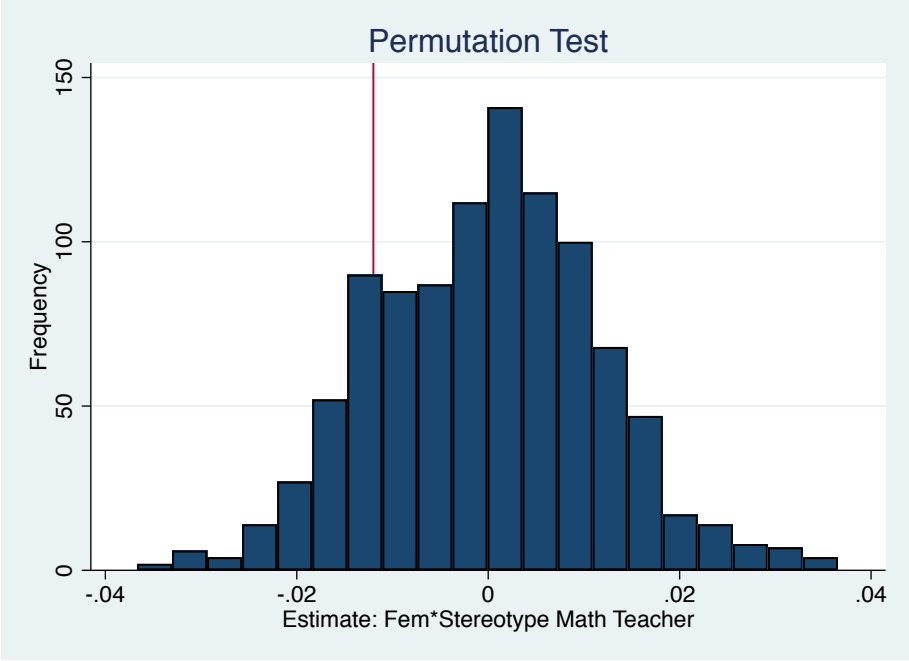
Notes: This graph shows the difference in math PISA test scores between females and males. Countries in which girls lag behind are colored in red, while countries where boys lag behind are colored in green.
Source: Author's calculation on PISA data (2015).

Figure A.V. Permutation Tests

Panel A

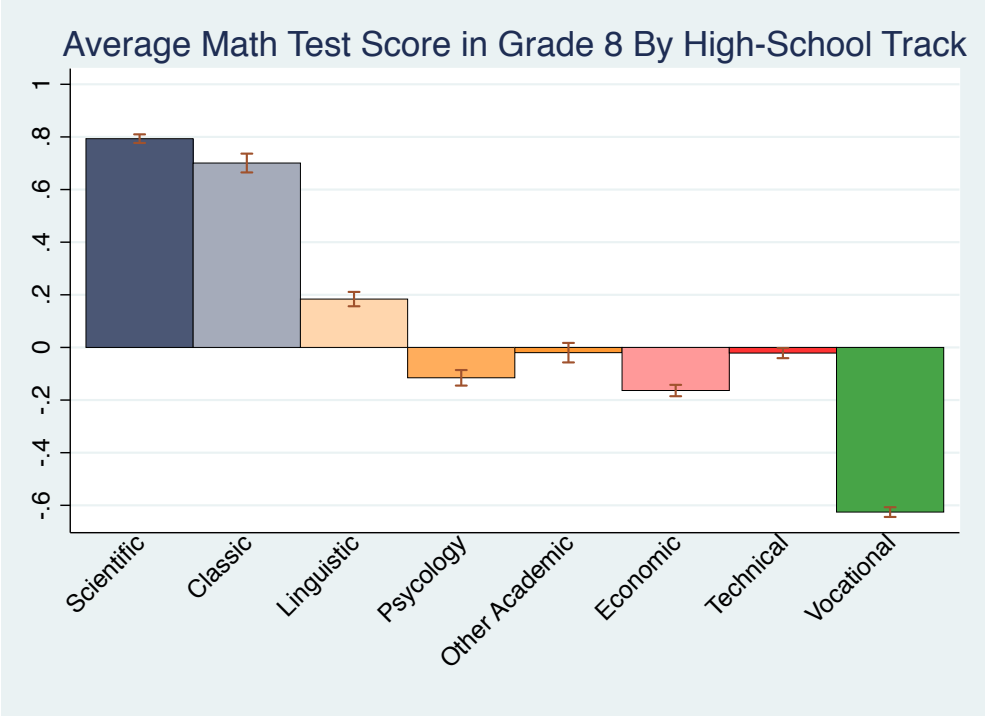


Panel B



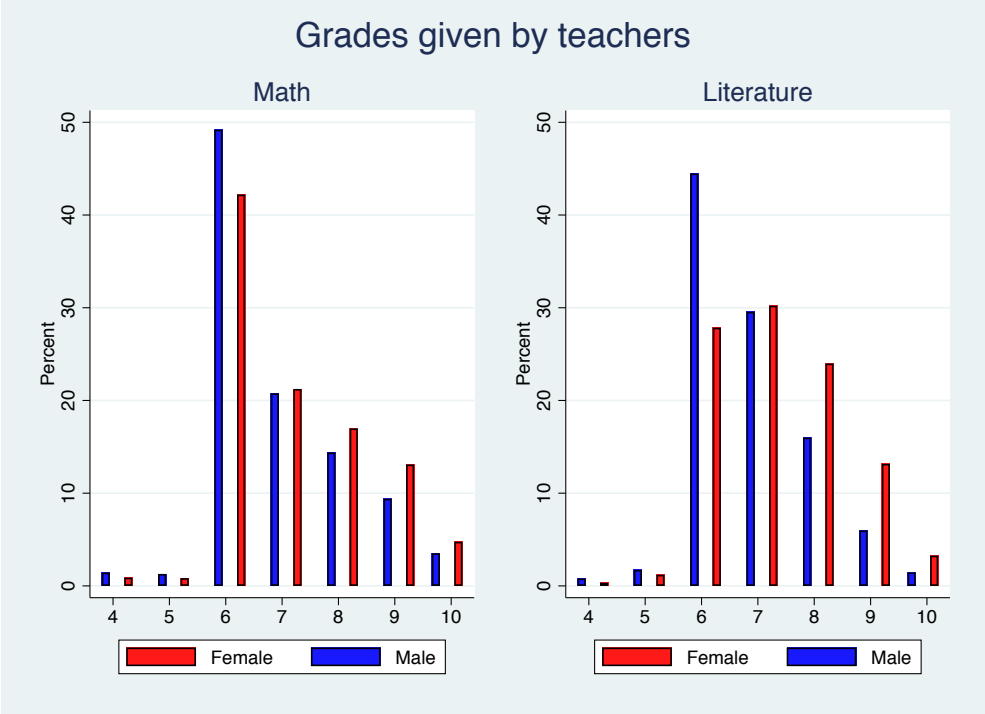
Notes: This figures plots the "Fem*Stereotypes" of a permutation test that runs the main regression in equation (1) 1000 times randomly assigning the stereotypes to math teachers in Panel A and literature teachers in Panel B. The vertical (red) line signifies the coefficient from the main regression in column 2 of Table V.

Figure A.VI. Average Standardized Test Score in Grade 8 by High-School Track



Notes: This graph shows the average standardized test score in grade 8 by high-school track selected.

Figure A.VII. Grades given by teachers



Notes: This graph shows the distribution of grades given by teachers at the end of the school year.

Table A.I. Balance Table: students in the sample vs. other students

Panel A: Comparison between students in the sample and other Italian students				
Variable	(1) Other Italian Students	(2) Sample	(3) Diff.	(4) Norm. Diff.
Female	0.410	0.455	0.002	0.064
	(0.492)	(0.498)	(0.003)	
Immigrant	0.096	0.217	0.118	0.238
	(0.295)	(0.412)	(0.011)***	
Test score grade 8	56.636	55.547	-0.824	-0.040
	(19.037)	(19.699)	(0.696)	
Mother: Less than Diploma	0.244	0.227	-0.047	-0.029
	(0.430)	(0.419)	(0.016)***	
Mother: Diploma	0.330	0.378	0.014	0.071
	(0.470)	(0.485)	(0.015)	
Mother: More than Diploma	0.095	0.138	0.035	0.093
	(0.294)	(0.344)	(0.017)**	
Father: Less than Diploma	0.282	0.263	-0.053	-0.031
	(0.450)	(0.440)	(0.018)***	
Father: Diploma	0.291	0.334	0.013	0.065
	(0.454)	(0.472)	(0.013)	
Father: More than Diploma	0.084	0.126	0.037	0.098
	(0.277)	(0.332)	(0.017)**	
Mother: Low Occupation	0.384	0.376	-0.052	-0.011
	(0.486)	(0.484)	(0.019)***	
Mother: Intermediate Occupation	0.220	0.289	0.050	0.113
	(0.414)	(0.453)	(0.012)***	
Mother: High Occupation	0.071	0.094	0.017	0.059
	(0.257)	(0.292)	(0.011)	
Father: Low Occupation	0.242	0.282	0.017	0.065
	(0.428)	(0.450)	(0.017)	
Father: Intermediate Occupation	0.267	0.291	-0.003	0.039
	(0.442)	(0.454)	(0.012)	
Father: High Occupation	0.143	0.162	0.004	0.037
	(0.350)	(0.368)	(0.017)	
Class size	22.151	22.526	0.379	0.080
	(3.829)	(2.668)	(0.173)**	
Observations	3,100,239	41,733	3,141,972	

Table A.I. Balance Table: students in the sample vs. other students *(cont.)*

Panel B: Comparison between students in the sample and other students in the same provinces				
	(1)	(2)	(3)	(4)
Variable	Other Italian Students	Sample	Diff.	Norm. Diff.
Female	0.407	0.455	0.003	0.069
	(0.491)	(0.498)	(0.003)	
Immigrant	0.134	0.217	0.081	0.155
	(0.341)	(0.412)	(0.011)***	
Test score grade 8	56.563	55.547	-1.177	-0.037
	(19.039)	(19.699)	(0.700)*	
Mother: Less than Diploma	0.194	0.227	0.012	0.056
	(0.396)	(0.419)	(0.017)	
Mother: Diploma	0.362	0.378	-0.026	0.023
	(0.481)	(0.485)	(0.016)*	
Mother: More than Diploma	0.119	0.138	0.006	0.039
	(0.324)	(0.344)	(0.017)	
Father: Less than Diploma	0.238	0.263	-0.002	0.040
	(0.426)	(0.440)	(0.018)	
Father: Diploma	0.317	0.334	-0.020	0.025
	(0.465)	(0.472)	(0.014)	
Father: More than Diploma	0.107	0.126	0.008	0.042
	(0.309)	(0.332)	(0.017)	
Mother: Low Occupation	0.319	0.376	0.023	0.085
	(0.466)	(0.484)	(0.019)	
Mother: Intermediate Occupation	0.273	0.289	-0.015	0.026
	(0.446)	(0.453)	(0.013)	
Mother: High Occupation	0.094	0.094	-0.012	-0.001
	(0.292)	(0.292)	(0.011)	
Father: Low Occupation	0.225	0.282	0.035	0.093
	(0.417)	(0.450)	(0.017)**	
Father: Intermediate Occupation	0.276	0.291	-0.017	0.023
	(0.447)	(0.454)	(0.012)	
Father: High Occupation	0.168	0.162	-0.028	-0.013
	(0.374)	(0.368)	(0.018)	
Class size	22.537	22.526	0.015	-0.003
	(3.162)	(2.668)	(0.180)	
Observations	418,408	41,733	460,141	

Notes: data from the standardized test score INVALSI of all Italian students in grade 8 from 2011-12 to 2016-17. 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 literature teacher. The normalized difference shown in column 4 is the formula recommended by Imbens and Wooldridge (2009). More details are reported in footnote 11 of the paper.

Table A.II. Balance table of the differences between teachers matched (not matched) with students graduating from 2012 to 2017

	Math Teachers				Literature Teachers			
	Not matched	Matched	Dif.	se	Not matched	Matched	Dif.	se
Female	0.771	0.811	-0.039	(0.047)	0.845	0.899	-0.055*	(0.024)
Born in the North	0.573	0.611	-0.038	(0.061)	0.547	0.719	-0.172***	(0.037)
Age	42.158	48.806	-6.648***	(1.182)	40.858	49.550	-8.692***	(0.688)
Full time contract	0.461	0.788	-0.327***	(0.053)	0.643	0.944	-0.302***	(0.025)
Yeas of experience	10.455	18.829	-8.375***	(1.457)	11.472	21.668	-10.196***	(0.795)
Teaching in 2015-16	0.415	0.869	-0.455***	(0.044)	0.444	0.952	-0.508***	(0.025)
Children	0.590	0.683	-0.092	(0.056)	0.504	0.706	-0.201***	(0.036)
Number of children	1.659	1.876	-0.217	(0.136)	1.573	1.781	-0.208*	(0.090)
Number of daughters	0.864	0.866	-0.002	(0.125)	0.909	0.845	0.064	(0.083)
Low edu Mother	0.452	0.518	-0.066	(0.063)	0.346	0.492	-0.145***	(0.040)
Middle edu Mother	0.384	0.331	0.053	(0.060)	0.434	0.373	0.061	(0.040)
High edu Mother	0.164	0.151	0.013	(0.046)	0.220	0.135	0.084**	(0.030)
Advanced STEM	0.282	0.231	0.051	(0.052)	0.000	0.000	0.000	(0.000)
Math Olympiad	0.051	0.163	-0.112*	(0.043)	0.000	0.000	0.000	(0.000)
Refresher courses	0.821	0.914	-0.094*	(0.037)	0.825	0.926	-0.101***	(0.023)
Degree Laude	0.250	0.209	0.041	(0.055)	0.372	0.326	0.046	(0.040)
IAT Gender	0.116	0.091	0.025	(0.046)	0.338	0.377	-0.039	(0.030)
IAT Race	0.461	0.465	-0.004	(0.031)	0.433	0.466	-0.033	(0.020)
Boys better in Invalsi	0.245	0.213	0.031	(0.063)	0.088	0.100	-0.011	(0.031)
Girls better in Invalsi	0.367	0.307	0.060	(0.071)	0.531	0.537	-0.006	(0.053)
Gender Equal in Invalsi	0.388	0.480	-0.092	(0.076)	0.381	0.363	0.018	(0.051)
Satisfy with teacher job	3.718	3.717	0.001	(0.109)	3.800	3.905	-0.105	(0.068)
WVS Gender Equality	0.167	0.158	0.009	(0.045)	0.112	0.105	0.006	(0.024)
Gender Dif Innate Ability	1.653	1.512	0.141	(0.095)	1.475	1.413	0.062	(0.055)
Observations	83		454		238	615		

Notes: First hand data from teachers' questionnaire. We compare teachers matched with students' data with those not matched. Notice that for 3 math and 9 literature teachers, there is no IAT test because they did not complete the test.

Table A.III. The impact of teacher characteristics on students' improvement in performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Std Math test score in grade 8						
Panel A: X=	**Teacher Female**		**Math Teachers Advanced STEM**		**Degree cum laude**	
X	0.154***	0.079**	-0.030	-0.063**	0.048	0.052*
	(0.051)	(0.033)	(0.058)	(0.031)	(0.048)	(0.027)
Fem*X	0.026	0.026	-0.020	-0.023	0.019	0.015
	(0.033)	(0.029)	(0.035)	(0.032)	(0.028)	(0.025)
Obs.	30359	30359	30359	30359	30359	30359
R^2	0.012	0.178	0.009	0.178	0.009	0.178
Dependent variable: Std Math test score in grade 8						
Panel B: X=	**Math Olympics**		**Math Teachers Full-time contract**		**High experience**	
X	0.050	0.074**	0.195***	0.159***	0.236***	0.135**
	(0.055)	(0.033)	(0.048)	(0.036)	(0.064)	(0.061)
Fem*X	-0.013	-0.024	-0.027	-0.030	-0.056	-0.079
	(0.034)	(0.030)	(0.040)	(0.033)	(0.058)	(0.050)
Obs.	30359	30359	30359	30359	30359	30359
R^2	0.009	0.178	0.012	0.179	0.012	0.178
Dependent variable: Std Reading test score in grade 8						
Panel C: X=	**Teacher Female**		**Literature Teachers Full-time contract**		**High experience**	
X	0.103**	0.076***	0.082	0.008	0.053	0.064*
	(0.051)	(0.029)	(0.081)	(0.059)	(0.054)	(0.037)
Fem*X	-0.048	-0.040	-0.024	-0.018	0.008	-0.025
	(0.040)	(0.035)	(0.072)	(0.056)	(0.048)	(0.041)
Obs.	29486	29486	29486	29486	29486	29486
R^2	0.014	0.213	0.014	0.213	0.015	0.213
Student controls	No	Yes	No	Yes	No	Yes
School FE	No	Yes	No	Yes	No	Yes

Notes: This table reports OLS estimates of the impact of math teachers' characteristics on improvements in math performance of their students. The unit of observation is the student. Standard errors (in parentheses) are robust and clustered at the teacher level. All columns include a dummy for the gender of the student ("Fem"). "High experience" means more than 15 years, " Mid experience" from 5 to 15. Student controls include parents' education and occupation and immigration status interacted with student gender. An intermediate level of experience and a missing category for experience (and the interactions with female) are included in columns 5-6 of Panel B and C. The level of this variable is statistically significant only for column 5 of Panel B. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.IV. Correlation between teachers' characteristics and Gender IAT Score

	Math Teachers		**Literature Teachers**	
	Matched sample	**All teachers**	**Matched sample**	**All teachers**
	(1)	(2)	(3)	(4)
Female	-0.166***	-0.184***	0.552***	0.568***
	(0.053)	(0.045)	(0.057)	(0.047)
Born in the North	-0.055	-0.066*	-0.054*	-0.041
	(0.038)	(0.034)	(0.030)	(0.025)
Age	-0.014	-0.028	0.000	-0.001
	(0.025)	(0.021)	(0.018)	(0.015)
Age sq.	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
High Edu Mother	0.031	0.013	-0.006	0.008
	(0.037)	(0.036)	(0.033)	(0.027)
Children	-0.003	0.024	-0.052	-0.046
	(0.059)	(0.055)	(0.046)	(0.039)
Daughters	0.054	0.017	0.092**	0.054
	(0.051)	(0.050)	(0.039)	(0.033)
Degree Laude	-0.093**	-0.060	-0.000	0.028
	(0.041)	(0.042)	(0.029)	(0.026)
Full time contract	-0.056	-0.004	0.017	0.056
	(0.071)	(0.057)	(0.067)	(0.049)
High Experience	-0.030	0.020	-0.098	-0.065
	(0.100)	(0.086)	(0.112)	(0.072)
Med Experience	0.052	0.087	-0.002	0.021
	(0.080)	(0.064)	(0.097)	(0.064)
WVS Gender Equality	-0.025	-0.009	0.041	0.053
	(0.052)	(0.048)	(0.045)	(0.035)
Gender Dif Innate Ability	0.031	0.027	-0.014	-0.001
	(0.024)	(0.020)	(0.024)	(0.020)
Advanced STEM	-0.094**	-0.093**		
	(0.044)	(0.042)		
Math Olympiad	0.081	0.074		
	(0.057)	(0.054)		
Obs.	454	534	615	844
R^2	0.105	0.101	0.223	0.249

Notes: This table reports OLS estimates of the correlation between literature teacher stereotypes measured by IAT score and own teacher characteristics. The unit of observation is teacher t in school s. Standard errors (in parentheses) are robust and clustered at the school level. The description of all variables is the same as in Table III. We include the order of IATs for math teachers (if the first one was the gender IAT and if the first associations were order compatible or not) and missing categories if the information is not available. The restricted sample includes data of teachers used in the main regressions of this paper. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.V. Exogeneity of assignment of students to math teachers with different bias

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A- Dependent Variable: Math Teacher implicit gender bias (standardized)						
Share Fem	0.055	0.048	0.004	-0.030	-0.125	0.589
	(0.934)	(0.925)	(0.942)	(0.904)	(0.871)	(1.350)
Share Male HighEduMother		-1.209**			-1.203*	-1.203
		(0.563)			(0.638)	(0.949)
Share Fem HighEduMother		0.255			0.202	-0.665
		(0.448)			(0.549)	(0.902)
Share Male MedOccFather			-0.074		0.428	-0.229
			(0.772)		(0.840)	(1.037)
Share Fem MedOccFather			-0.026		-0.187	0.413
			(0.623)		(0.796)	(1.098)
Share Male HighOccFather			0.010		0.436	1.496
			(0.978)		(0.965)	(1.355)
Share Fem HighOccFather			-0.830		-0.817	0.031
			(0.834)		(1.002)	(1.615)
Share Male Immigrant				0.447	0.574	0.912
				(0.618)	(0.631)	(0.927)
Share Fem Immigrant				-0.464	-0.758	-0.703
				(0.620)	(0.710)	(0.864)
Male Average Std Math 5						-0.084
						(0.259)
Fem Average Std Math 5						-0.180
						(0.212)
Obs.	454	454	454	454	454	282
R^2	0.202	0.215	0.207	0.205	0.226	0.239

Table A.V. Exogeneity of assignment of students to math teachers with different bias

	(1)	(2)	(3)	(4)	(5)	(6)
Panel B- Dependent Variable: Literature Teacher implicit gender bias (standardized)						
Share Fem	-0.367	-0.387	-0.328	-0.395	-0.384	0.009
	(0.594)	(0.607)	(0.603)	(0.610)	(0.626)	(0.843)
Share Male HighEduMother		0.005			0.064	0.077
		(0.416)			(0.429)	(0.653)
Share Fem HighEduMother		-0.306			-0.359	-0.972*
		(0.335)			(0.433)	(0.509)
Share Male MedOccFather			-0.889*		-0.981**	-1.139
			(0.462)		(0.491)	(0.759)
Share Fem MedOccFather			0.052		0.083	0.575
			(0.425)		(0.449)	(0.597)
Share Male HighOccFather			-0.095		-0.208	0.359
			(0.555)		(0.594)	(0.822)
Share Fem HighOccFather			0.031		0.205	0.271
			(0.527)		(0.622)	(0.994)
Share Male Immigrant				-0.394	-0.497	-0.874
				(0.424)	(0.452)	(0.551)
Share Fem Immigrant				0.009	-0.028	0.104
				(0.363)	(0.391)	(0.523)
Male Average Std Math 5						-0.126
						(0.161)
Fem Average Std Math 5						0.090
						(0.133)
Obs.	616	616	616	616	616	357
R^2	0.177	0.181	0.184	0.179	0.191	0.233

Notes: This table reports OLS estimates of the correlation between teacher stereotypes measured by IAT score and student characteristics. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the school level. The number of clusters is 98 in columns 1-6 and 73 in column 7. All regressions include controls for the order of IAT in the questionnaire administered. The last column has a lower number of observations since the test score in grade 5 is available only for part of the sample. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VI. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

	All		Female Teachers		Male Teachers	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dep. Variable - Std Math Grade 8						
Female	-0.198***	-0.033	-0.196***	0.002	-0.203***	-0.193
	(0.013)	(0.082)	(0.015)	(0.094)	(0.029)	(0.202)
Fem*Teacher Stereotypes	-0.032**	-0.031***	-0.030**	-0.031**	-0.040	-0.059**
	(0.013)	(0.011)	(0.014)	(0.012)	(0.029)	(0.028)
Obs.	30359	30359	25550	25550	4809	4809
R^2	0.206	0.276	0.205	0.275	0.193	0.269
Panel B: Dep. Variable - Std Reading Grade 8						
Female	0.225***	0.308***	0.220***	0.266***	0.258***	0.366
	(0.013)	(0.093)	(0.014)	(0.094)	(0.050)	(0.265)
Fem*Teacher Stereotypes	-0.012	-0.001	-0.007	0.001	-0.011	-0.014
	(0.013)	(0.013)	(0.015)	(0.014)	(0.039)	(0.046)
Obs.	29484	29484	26901	26901	2583	2583
R^2	0.181	0.291	0.181	0.291	0.179	0.291
Student controls	No	Yes	No	Yes	No	Yes
Teacher controls	No	Yes	No	Yes	No	Yes

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math and reading standardized test score in grade 8 in Panel A and B, respectively. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VII. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression

	Dep. Variable: Std Math 8th grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.154***	-0.131***	0.005	-0.148***	-0.122***	0.014
	(0.018)	(0.028)	(0.084)	(0.021)	(0.030)	(0.085)
Fem *No Bias				-0.051	-0.054*	-0.051*
				(0.032)	(0.030)	(0.029)
Fem *Male-Math				-0.044	-0.049*	-0.047*
				(0.027)	(0.026)	(0.025)
Fem *IAT>0 Male-Math	-0.049**	-0.048**	-0.048**			
	(0.024)	(0.022)	(0.022)			
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Obs.	30359	30359	30359	30359	30359	30359
R^2	0.206	0.275	0.276	0.206	0.275	0.276

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student *i* in class *c* taught by teacher *t* in grade 8 of school *s*. Instead of using a continuous variable as teacher bias, I use categorical variables. In columns 1-3, we consider a positive or negative sign in the IAT score. In columns 4-6, we consider the thresholds defined by Greenwald et al. (2003) where no bias is the interval of IAT raw score between -0.15 and +0.15. Standard errors (in parentheses) are robust and clustered at the teacher level. The number of clusters is 454. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.VIII. Estimation of the effect of teachers' gender stereotypes

Dependent Variable: Math standardized test score in grade 8						
Heterogeneous effects by	**Student Characteristics**				**Interaction time with teacher**	
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.032	0.089	-0.026	-0.032	-0.037	-0.015
	(0.082)	(0.120)	(0.082)	(0.082)	(0.082)	(0.085)
Fem*Teacher Stereotypes(IAT)	-0.031***	-0.096***	-0.019	-0.030**	-0.029**	-0.037***
	(0.011)	(0.027)	(0.020)	(0.013)	(0.012)	(0.013)
Fem*IAT*Top tercile Math6		0.136***				
		(0.043)				
Fem*IAT*Middle tercile Math6		0.024				
		(0.039)				
Fem*IAT*HighEduM			-0.000			
			(0.026)			
Fem*IAT*Missing EduM			-0.051			
			(0.031)			
Fem*IAT*Immigrant				-0.008		
				(0.030)		
Fem*IAT*Extended School Day					-0.009	
					(0.029)	
Fem*IAT*New Math Teacher						0.031
						(0.031)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	30359	9540	30359	30359	30359	30359
R^2	0.276	0.575	0.276	0.276	0.276	0.276

Notes: This table reports OLS estimates of the heterogeneous impact of math teachers' gender stereotypes measured by IAT score on math standardized test score in grade 8 by observable characteristics of the student and by interaction time with teacher. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level. The number of clusters is 454. The number of observations is lower in the last column because for the first two cohorts of students who started middle school before 2011, I do not have information on who their teacher was in grade 6 and/or 7. The variable "Fem" indicates the gender of the student, "HighEduM" whether the mother has at least a diploma, "Missing EduM" whether the information on education of the mother is missing, "tercile Math6" is the tercile of standardized test score in math in grade 6, and "Immigrant" indicates whether the student is not Italian citizen. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. Regressions are all fully saturated even if not all interactions are shown in the table. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.IX. Estimation of the effect of teachers' gender stereotypes for one year on math standardized test score in grade 6 and 8 - class FE regression

	Std Math 6th grade			**Std Math 8th grade**		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.209***	-0.202***	0.140	-0.190***	-0.192***	0.033
	(0.029)	(0.027)	(0.227)	(0.032)	(0.030)	(0.171)
Fem*Math Teacher Stereotypes	0.004	0.005	0.010	0.011	-0.006	0.010
	(0.033)	(0.030)	(0.030)	(0.028)	(0.027)	(0.029)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	No	Yes	Yes
Student Controls*Fem	No	No	Yes	No	No	Yes
Teacher Controls	No	No	Yes	No	No	Yes
Obs.	5532	5532	5532	4695	4695	4695
R^2	0.214	0.275	0.278	0.227	0.298	0.302

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student i in class c taught by teacher t in grade 6 in columns 1-3 and grade 8 in columns 4-6 of school s. The sample is restricted to teachers teaching in grade 6 for columns 1-3 and for teachers assigned to students only in grade 8 for columns 4-6. Standard errors (in parentheses) are robust and clustered at the teacher level . The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.X. Estimation of the effect of teachers' gender stereotypes on the probability of missing standardized test score in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)
Panel A- Dependent Variable: Missing math standardized test score in grade 8					
Female	-0.025***	-0.023***	-0.014***	-0.006*	0.023
	(0.003)	(0.003)	(0.003)	(0.004)	(0.017)
Fem*Math Teacher Stereotypes		0.005	0.003	0.002	0.002
		(0.003)	(0.003)	(0.003)	(0.003)
Fem* Teacher Fem					0.010
					(0.006)
Fem* Teacher. Born North					-0.004
					(0.005)
Fem*Advanced STEM Teacher					-0.005
					(0.006)
Obs.	32642	32642	32642	32642	32642
R^2	0.102	0.102	0.377	0.380	0.380
Panel B- Dependent Variable: Missing reading standardized test score in grade 8					
Female	-0.024***	-0.023***	-0.012***	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.021)
Fem*Lit Teacher Stereotypes		-0.004	-0.002	-0.003	-0.003
		(0.003)	(0.003)	(0.003)	(0.003)
Fem* Teacher Fem					0.005
					(0.010)
Fem* Teacher. Born North					-0.011*
					(0.006)
Obs.	31701	31701	31701	31701	31701
R^2	0.089	0.089	0.367	0.368	0.368
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1, where the dependent variable is the probability of not taking the math or reading standardized test score in grade 8 in Panel A and B, respectively. The data may be missing because the student did not take the test or because the school did not correctly match the test score with the administrative registry. The dependent variables refer to math teachers in Panel A and to literature teachers in Panel B. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level (454 in Panel A and 615 in Panel B). The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

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

Dependent Variable: Math standardized test score in grade 8			
	(1)	(2)	(3)
Graduation	All	2012-2016	2017
Female	-0.032	-0.026	-0.030
	(0.082)	(0.101)	(0.142)
Fem*Teachers' Stereotypes	-0.031***	-0.029**	-0.037*
	(0.011)	(0.014)	(0.021)
Class FE	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes
Obs.	30359	22657	7702
R^2	0.276	0.278	0.275

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XII. Estimation of the effect of teachers' gender stereotypes on math standardized test score in grade 8 - class FE regression -sample restricted to cohorts by schools where students are randomly assigned according to all characteristics by gender

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent Variable - Math standardized test score in grade 8					
Female	-0.175***	-0.189***	-0.190***	-0.158***	-0.116
	(0.016)	(0.019)	(0.018)	(0.035)	(0.109)
Fem*Math Teacher Stereotypes		-0.031*	-0.033*	-0.033*	-0.035**
		(0.018)	(0.017)	(0.017)	(0.018)
Fem*Math Teacher Fem					0.058
					(0.041)
Fem*North Math Teacher					0.007
					(0.032)
Fem*Advanced STEM Teacher					0.004
					(0.035)
Obs.	15781	15781	15781	15781	15781
R^2	0.201	0.201	0.272	0.273	0.273
Panel B: Dependent Variable - Reading standardized test score in grade 8					
Female	0.226***	0.230***	0.228***	0.245***	0.430***
	(0.016)	(0.017)	(0.016)	(0.033)	(0.135)
Fem*Literature Teacher Stereotypes		-0.012	-0.010	-0.011	-0.009
		(0.017)	(0.016)	(0.016)	(0.016)
Fem*Literature Teacher Fem					-0.032
					(0.059)
Fem*North Literature Teacher					0.002
					(0.039)
Obs.	15916	15916	15916	15916	15916
R^2	0.172	0.172	0.281	0.281	0.282
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math standardized test score in grade 8. I restrict the sample to schools by cohorts where Pearson Chi-Square tests suggest statistically independence of all student characteristics (gender, education of the mother, occupation of the father, immigrant dummy, generation of immigration) and of all student characteristics by gender. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XIII. Estimation of the effect of teachers' gender stereotypes on math and reading standardized test score in grade 8 - class FE regression

Dependent Variable:	**Std math 8th grade**				**Std reading 8th grade**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fem	-0.176***	-0.179***	-0.193***	-0.204	0.219***	0.214***	0.222***	0.392***
	(0.015)	(0.017)	(0.018)	(0.134)	(0.015)	(0.017)	(0.018)	(0.105)
Fem*Lit Teach Stereotypes		0.008	0.006	0.022			-0.022	-0.016
		(0.017)	(0.017)	(0.017)			(0.017)	(0.016)
Fem*Math Teach Stereotypes			-0.033**	-0.029*		-0.010	-0.011	-0.014
			(0.016)	(0.015)		(0.017)	(0.017)	(0.016)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	No	Yes	No	No	No	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes
Obs.	18102	18102	18102	18102	18102	18102	18102	18102
R^2	0.198	0.198	0.198	0.274	0.183	0.183	0.183	0.299

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math or reading standardized test score in grade 8. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. I restrict the sample to classes for which I have information on the implicit associations of both literature and math teachers. Standard errors (in parentheses) are robust and clustered at the class level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between students' gender and teacher characteristics. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XIV. Correlation between track choice and recommendation

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var	**Vocational**			**Scientific**		
Vocational recommendation	0.327***	0.288***	0.280***			
	(0.010)	(0.011)	(0.011)			
Fem*Vocational recommendation	0.101***	0.094***	0.093***			
	(0.014)	(0.014)	(0.015)			
Scientific recommendation				0.670***	0.582***	0.561***
				(0.011)	(0.012)	(0.013)
Fem*Scientific recommendation				-0.015	-0.006	0.020
				(0.015)	(0.015)	(0.015)
Female	0.004	-0.003	0.016	-0.061***	-0.051***	-0.023**
	(0.003)	(0.004)	(0.012)	(0.006)	(0.006)	(0.010)
Constant	0.036***	0.043***	0.067***	0.162***	0.152***	0.096***
	(0.003)	(0.004)	(0.009)	(0.004)	(0.004)	(0.008)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Sq. Test score 8	No	Yes	Yes	No	Yes	Yes
Student controls	No	No	Yes	No	No	Yes
Student controls*Fem	No	No	Yes	No	No	Yes
Obs.	20646	20646	20646	20646	20646	20646
R^2	0.355	0.363	0.367	0.486	0.507	0.513

Notes: This table reports OLS estimates of the correlation between track choice and teachers' track recommendation, for vocational in columns 1-3 and scientific track in columns 4-6. The dataset used in this table is the same as in the main analysis of the paper, but here I restrict the sample to students for whom I have information on track choice, teachers' recommendation, and test score in grade 8. Standard errors (in parentheses) are robust and clustered at the class level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student in columns 3 and 6. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XV. Estimation of the effect of teachers' gender stereotypes on high-school track choice - school by cohort FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A- Dependent Variable: Vocational High-School Track Choice								
Fem	0.011**	0.019***	0.019***	0.063	0.052	0.043	0.025***	0.009
	(0.006)	(0.006)	(0.006)	(0.046)	(0.045)	(0.043)	(0.009)	(0.006)
Fem*Math Teach Stereotypes		0.018***	0.016**	0.014**	0.009	0.005	0.023**	
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	
Math Teach Stereotypes		-0.008*	-0.008*	-0.008	-0.008	-0.007	-0.002	
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	
Fem*Literature Teach Stereotypes							0.010	-0.000
							(0.009)	(0.006)
Literature Teach Stereotypes							-0.002	-0.000
							(0.006)	(0.005)
Constant	0.163***	0.160***	0.186***	0.172***	0.198***	0.145***	0.153***	0.162***
	(0.004)	(0.004)	(0.008)	(0.034)	(0.035)	(0.034)	(0.006)	(0.004)
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R^2	0.067	0.068	0.116	0.120	0.108	0.176	0.078	0.068
Effect on girls		0.010**	0.007	0.006	0.002	-0.002	0.021***	
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	
Panel B- Dependent Variable: Scientific High-School Track Choice								
Fem	-0.089***	-0.093***	-0.091***	-0.029	-0.040	-0.033	-0.106***	-0.092***
	(0.007)	(0.007)	(0.007)	(0.048)	(0.051)	(0.049)	(0.012)	(0.007)
Fem*Math Teach Stereotypes		-0.010	-0.007	-0.008	-0.009	-0.007	-0.021**	
		(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.010)	
Math Teach Stereotypes		0.003	0.004	0.002	0.000	0.002	0.016**	
		(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.008)	
Fem*Literature Teach Stereotypes							0.002	-0.003
							(0.010)	(0.007)
Literature Teach Stereotypes							-0.001	0.002
							(0.008)	(0.006)
Constant	0.283***	0.285***	0.193***	0.121***	0.119***	0.136***	0.289***	0.283***
	(0.005)	(0.005)	(0.008)	(0.040)	(0.043)	(0.043)	(0.008)	(0.005)
Obs.	21015	21015	21015	21015	19506	19506	11302	20254
R^2	0.068	0.068	0.109	0.118	0.117	0.229	0.076	0.068
Effect on girls		-0.006	-0.003	-0.007	-0.008	-0.005	-0.005	
		(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	
School cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	Yes	Yes	No	No

Notes: This table reports OLS estimates of equation 2, where the dependent variable is the high-school track choice. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVI. Estimation of the effect of teachers' gender stereotypes on high-school track recommendation of teachers - school by cohort FE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A- Dependent Variable: Vocational High-School Track Recommendation								
Fem	-0.077***	-0.070***	-0.064***	-0.077	-0.090	-0.111**	-0.058***	-0.086***
	(0.008)	(0.009)	(0.008)	(0.060)	(0.062)	(0.055)	(0.013)	(0.009)
Fem*Math Teach Stereotypes		0.018**	0.013	0.013	0.012	0.004	0.020*	
		(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.011)	
Math Teach Stereotypes		-0.007	-0.008	-0.011**	-0.013**	-0.014**	-0.000	
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	
Fem*Literature Teach Stereotypes							0.003	0.007
							(0.012)	(0.009)
Literature Teach Stereotypes							-0.003	-0.011
							(0.009)	(0.006)
Constant	0.435***	0.432***	0.516***	0.520***	0.540***	0.382***	0.426***	0.434***
	(0.005)	(0.006)	(0.010)	(0.046)	(0.048)	(0.045)	(0.008)	(0.006)
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R^2	0.093	0.094	0.231	0.234	0.203	0.390	0.092	0.089
Effect on girls		0.011*	0.005	0.001	-0.001	-0.009	0.019**	
		(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	
Panel B- Dependent Variable: Scientific High-School Track Recommendation								
Fem	-0.029***	-0.033***	-0.035***	-0.057	-0.060	-0.052	-0.042***	-0.032***
	(0.006)	(0.006)	(0.006)	(0.045)	(0.048)	(0.045)	(0.010)	(0.007)
Fem*Math Teach Stereotypes		-0.008	-0.006	-0.009	-0.010	-0.007	-0.015	
		(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	
Math Teach Stereotypes		-0.001	-0.001	0.000	-0.000	0.000	0.000	
		(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)	
Fem*Literature Teach Stereotypes							0.003	-0.002
							(0.010)	(0.007)
Literature Teach Stereotypes							-0.010	-0.006
							(0.008)	(0.006)
Constant	0.196***	0.195***	0.121***	0.096***	0.089**	0.122***	0.201***	0.204***
	(0.004)	(0.005)	(0.006)	(0.036)	(0.040)	(0.038)	(0.007)	(0.005)
Obs.	18684	18684	18684	18684	16904	16904	10575	17979
R^2	0.174	0.174	0.225	0.229	0.229	0.353	0.151	0.158
Effect on girls		-0.010*	-0.006	-0.009*	-0.010*	-0.007	-0.014*	
		(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	
School cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes	Yes	No	No
Student Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Teacher Controls*Fem	No	No	No	Yes	Yes	Yes	No	No
Sq. Math Test 8	No	No	No	No	Yes	Yes	No	No

Notes: This table reports OLS estimates of equation 2, where the dependent variable is the high-school track recommendation of teachers. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the class level. Columns 5 and 6 restrict the sample only to those students who took the standardized test in grade 8, while column 7 includes only students for whom we have information about both the math and literature teacher. The variable "Fem" indicates the gender of the student and "Stereotypes" is the IAT score of the teacher. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVII. Estimation of the effect of teachers' explicit and implicit bias on standardized test score in grade 8 - class FE regression

Dependent Variable	Std Math 8th grade				Std Reading 8th grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.130***	-0.102***	0.028	0.017	0.214***	0.221***	0.397***	0.307***
	(0.027)	(0.033)	(0.085)	(0.086)	(0.040)	(0.042)	(0.097)	(0.099)
Fem*Dif InnateAbility(MathTeach)	-0.060**	-0.065**	-0.067**	-0.069**				
	(0.031)	(0.027)	(0.028)	(0.028)				
Fem*Math Teacher Stereotypes				-0.032***				
				(0.011)				
Fem*Dif InnateAbility(LitTeach)					0.009	0.004	0.016	0.001
					(0.042)	(0.039)	(0.038)	(0.038)
Fem*Lit Teacher Stereotypes								-0.001
								(0.013)
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Teacher Controls	No	No	Yes	Yes	No	No	Yes	Yes
Obs.	30359	30359	30359	30359	29486	29486	29486	29486
R^2	0.206	0.275	0.276	0.276	0.181	0.291	0.291	0.291

Notes: This table reports OLS estimates, where the dependent variable is the standardized test score in grade 8. The unit of observation is student *i* in class *c* taught by teacher *t* in grade 8 of school *s*. Standard errors (in parentheses) are robust and clustered at the teacher level . The variable "Fem" indicates the gender of the student, "Dif Innate Ability" assumes value 1 if the teacher thinks there are innate differences in math abilities between men and women, 0 if there are few or no differences. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree with honour, refresher courses, age, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

Table A.XVIII. Estimation of the effect of teachers' gender stereotypes on grading by teacher

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A- Dependent Variable	**Math Grade end middle school**				**High Math Grade (>than6)**			
Fem	0.391***	0.389***	0.348***	0.348***	0.130***	0.130***	0.097**	0.097**
	(0.015)	(0.016)	(0.101)	(0.101)	(0.006)	(0.007)	(0.046)	(0.046)
Fem* Math Teacher Stereotypes		-0.005	-0.006	-0.006		-0.001	-0.002	-0.002
		(0.018)	(0.018)	(0.018)		(0.008)	(0.008)	(0.008)
Constant	6.662***	6.662***	6.563***	6.563***	0.460***	0.460***	0.408***	0.408***
	(0.010)	(0.010)	(0.024)	(0.024)	(0.004)	(0.004)	(0.011)	(0.011)
Obs.	21424	21424	21424	21424	21424	21424	21424	21424
R^2	0.523	0.523	0.532	0.532	0.398	0.398	0.409	0.409
Panel B- Dependent Variable	**Literature Grade end middle school**				**High Literature Grade (>than6)**			
Female	0.335***	0.336***	0.314***	0.564***	0.124***	0.125***	0.156***	0.294***
	(0.013)	(0.014)	(0.027)	(0.147)	(0.006)	(0.007)	(0.014)	(0.070)
Fem*Lit Teacher Stereotypes		-0.003	-0.003	0.005		-0.003	-0.003	-0.002
		(0.013)	(0.013)	(0.015)		(0.006)	(0.006)	(0.007)
Constant	6.752***	6.752***	6.641***	6.640***	0.562***	0.562***	0.502***	0.502***
	(0.008)	(0.008)	(0.021)	(0.021)	(0.004)	(0.004)	(0.011)	(0.011)
Obs.	20514	20514	20514	20514	20514	20514	20514	20514
R^2	0.572	0.572	0.583	0.584	0.412	0.412	0.425	0.425
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sq Std Test 8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	No	No	Yes	Yes
Teacher Controls	No	No	No	Yes	No	No	No	Yes

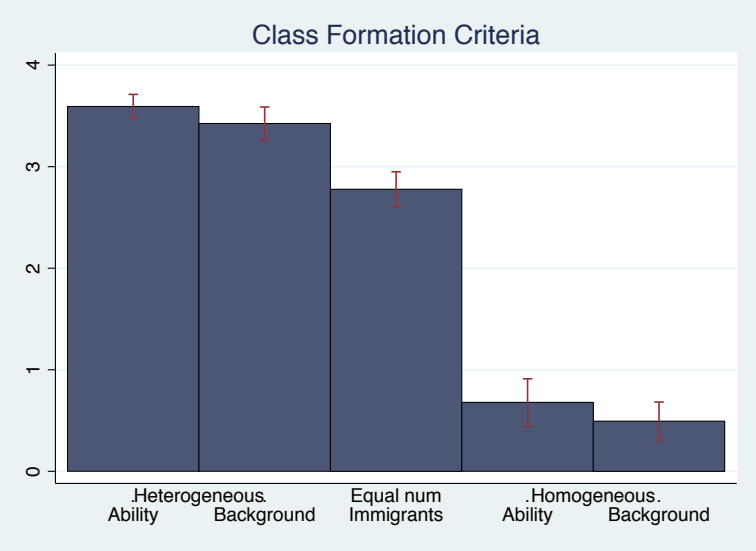
Notes: This table reports OLS estimates of equation 1, where the dependent variable is the grades given by teachers in columns 1-4 and a dummy for a grade higher than 7 in column 4-8. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level . The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, age, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

B Survey to School Principals

School principals were asked to complete a paper questionnaire, including information about the career counseling service offered by the school to students, class formation criteria at the beginning of middle school and refresher courses offered to teachers. 82 principals completed the questionnaire. Among them, 94% have a full-time contract, the average experience is 8.2 years, and in 15% of schools the principal is mainly ruling a different complex of school and has been assigned the direction of the institution in our sample temporarily until new staff is hired. This practice is widespread in Italy.

Among 81 principals who completed the questions on class formation, 64 percent consider ability heterogeneity within classes "Extremely important" and 33 percent "Important". Socioeconomic heterogeneity within classes is considered "Extremely important" by 60 percent of principals and "Important" by 29 percent. The equal allocation of immigrants across classes is considered "Extremely Important" by 25 percent of principals and "Important" by 38 percent. The summary statistics on the importance given to the different criteria in class formation are summarized in Figure B.I.

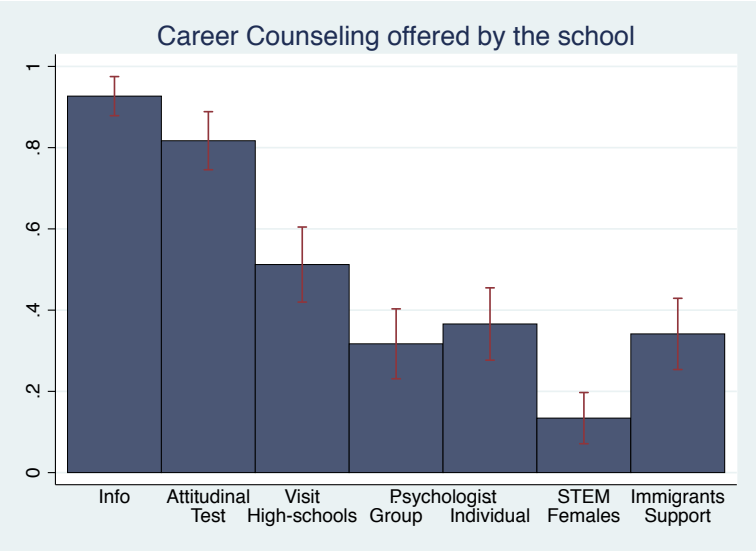
Figure B.I. Class formation criteria according to principals



An entire section of the questionnaire is dedicated to career counseling practices done in the school and the share of schools offering the different services is reported in Figure B.II. Most institutions claim to offer information on high-school curricula and attitudinal tests that help high-school choice. Around one third of schools organize meetings with psychologists at the

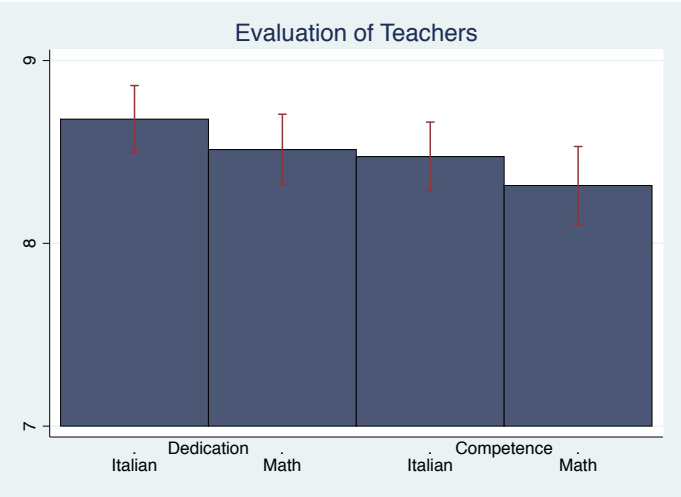
individual or group level to induce students to reflect on this important choice. However, very few institutions try to direct females toward STEM education.

Figure B.II. Share of schools offering different career counseling services



Principals are asked to assign a score from 1 to 10 to the dedication and competence of math and literature teachers in the middle school: generally principals are generous with the evaluation, but, if anything, math teachers are considered marginally less committed to their work and less competent than their literature colleagues, as shown in Figure B.III.

Figure B.III. Dedication and competence of teachers according with principals



C Teacher Survey

C.1 Gender Implicit Association Test

The concept behind IAT is that the easier the mental task, the faster the response production and the fewer the errors made in the process.[1]

I invite teachers to complete a seven-block IAT following the schematic overview presented in Figure C.I. Half of the teachers, randomly selected at the individual level completed the IAT as presented in Figure C.I, while the other half completed the IAT with the blocks in the following order: 1, 5, 6, 7, 2, 3, 4 ("order incompatible" IAT). The order of the two tasks is randomly selected at individual level. The blocks used to calculate the IAT score are blocks 3, 4, 6, and 7. The number of words that need to be categorized is 20 in blocks 3 and 6, and 40 in blocks 4 and 7, as in the standard IAT 7-blocks. The measure of implicit stereotypes is calculated as the difference in reaction time between the task in which scientific fields and male names are on the same side of the screen and the task in which scientific fields and female names are on the same side of the screen. The scoring procedure follows the guidelines of the improved scoring algorithm defined by Greenwald et al. (2003).

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

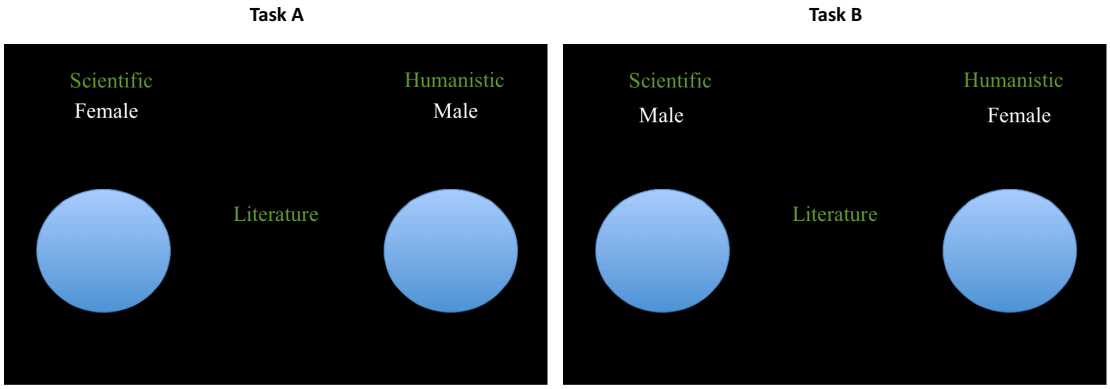
Figure C.I. Schematic overview of the Gender Implicit Association Test

Stimuli presented within each category are summarized in Figure C.II, while a screenshot of the tablet is shown in Figure C.III.

[1] It was initially developed by Donders (1868). Donders was very optimistic about the possibility of quantifying how the mind works using the "time required for simple mental processes" and performed some of the first experiments making participants respond with the right hand to stimuli on the right side and with the left hand to stimuli on the left side.

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

Figure C.II. Category Labels and Stimuli for the Implicit Association Tests



IAT score=0	No Bias	Time/Errors of Task A = Time/Errors of Task B
IAT score>0	Bias Pro Boys	Time/Errors of Task A > Time/Errors of Task B
IAT score<0	Bias Pro Girls	Time/Errors of Task A < Time/Errors of Task B

Figure C.III. Screenshot of the Implicit Association Test for the two types of tasks

Teachers were asked to complete a race IAT with male names and female names of Italians and immigrants. The order of race and gender IATs was randomized at the individual level. In Table C.I, I check the influence of order of blocks on the IAT score. On average, there is a small difference in the IAT score between individuals who perform the order compatible and incompatible test. Hence, in all regressions where there are no class (and therefore teacher) fixed effects, I control for the order of IATs.

Table C.I. Correlation between Gender-Science IAT score and order of different parts of the survey

Dependent variable : IAT score Math Teachers					
	(1)	(2)	(3)	(4)	(5)
First IAT Gender	0.048**			0.049**	0.050**
	(0.023)			(0.023)	(0.025)
First Questionnaire, then IAT		-0.093**		-0.096**	-0.078
		(0.045)		(0.046)	(0.055)
Order Compatible IAT Gender			-0.066***	-0.069***	-0.069***
			(0.021)	(0.021)	(0.022)
School FE	No	No	No	No	Yes
Obs.	1378	1378	1378	1378	1378
R^2	0.108	0.106	0.111	0.116	0.170

Notes: This table reports OLS estimates of the correlation between order of IAT and IAT score. A higher value of IAT score means stronger implicit association between Male-Science and Female-Literature. The dummy "First IAT Gender" captures the order of IATs (gender and race). The variable "Order Compatible IAT Gender" captures whether a teacher was asked first to associate compatible categories (Male-Scientific vs. Female-Humanistic) or the opposite (Female-Scientific vs. Male-Humanistic). Finally, in 8 cases for math teachers and 39 cases for literature teachers, I asked them to first complete a questionnaire and then the IATs. All columns include a dummy variable which assumes value 1 for math teachers. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

C.2 Teachers' Questionnaire

Factors influencing track choice

"Female students with the same math grade of males are less likely to attend a scientific track during high-school. According with your experience, how much can these factors influence the choice of females toward alternative tracks?" Answer in a scale from 1 to 5, where 1 means '*"Not at all"* and 5 means *"'A lot'*.

1. Low interest for scientific subjects
2. Low inclination for scientific subjects
3. Low self-esteem
4. Encouragement of the family toward alternative paths
5. Influence of gender predicament ("women are bad at math")

Factors influencing grading

"When you grade your students, which weight do you assign to the following components?" Answer in a scale from 1 to 5, where 1 means '*"A little"* and 5 means *"'A lot'*.

1. Performance in written exams
2. Performance in oral exams
3. Diligence in doing homeworks

Factors influencing track recommendation

"When you give the high-school track recommendation to your students, which weight do you assign to the following components?" Answer in a scale from 1 to 5, where 1 means '*"A little"* and 5 means *"'A lot'*.

1. Grades and performance at school
2. Predisposition and interests of the student
3. Parents' education
4. Economic resources of the family
5. Engagement of family in schooling

Teachers' Track recommendation

"When you give the high-school track recommendation, how much weight do you assign to these aspects?" Answer in a scale from 1 to 5, where 1 means '*"Not at all"* and 5 means *"'A lot'*.

- Performance at school
- Attitude and interest of students
- Education of parents
- Economic resources of families
- Participation and involvement of parents
- Presence of specific programs for immigrants (e.g. Italian courses)

Explicit gender bias

Do you agree with the following sentences?

- *There are innate biological differences in math abilities of women and men*: '*Not at all*", "*A little*", "*A bit*","*A lot*", "*Absolutely*"
- *When jobs are scarce, men should have more right to a job than women*: '*Agree*", "*Neither Agree nor Disagree*", "*Disagree*"

C.3 Students' Questionnaire

"Evaluate for each school subject your ability to tackle it.": '*Good*", "*Mediocre*", "*Scarce*"

D Effort vs. Talent

Research in social psychology shows that school teachers with higher gender stereotypes believe that males are more talented than females, even among students with comparable level of ability (Tiedemann, 2002). The performance of females is mainly attributed to effort. Furthermore, teachers are more likely to convey their stereotyping of *mathematics as a male domain* through their classroom instruction (Keller, 2001; Riegle-Crumb and Humphries, 2012) compared to *literature as a female domain*.

In this Appendix, I provide evidence that teacher perception of gender differences in effort and ability mirrors (or is mirrored by) pupils own perception. I have information on a sample of around 17000 Italian students in grade 6 collected by the Italian National Evaluation Center (INVALSI) on the reasons why students believe they are performing well in both math and reading. In particular, students are asked: "You correctly solved all exercises in a math exam. Why does it happen?" and "You present to your classmates a text you have read. You do it in a clear and precise way and everybody follows it with interest. Why does it happen?" Students can choose among different potential answers: "I was helped", "I was lucky, "It was easy", "I am good" and "I put in a lot of effort". Here I focus on the latter two answers. Female students are 7.6 and 2.3 percentage points less likely to say "I am good" in math and reading respectively than males (column 1 and 2, Table D.1). The difference in under- confidence between the two subjects is statistically significant for females, even controlling for individual level fixed effects (column 3). Finally, in column 4, I show that females are more likely than males to believe the most important reason of their success is effort. However, the difference between subjects is slightly lower, although statistically significant at 1 percent level. Unfortunately, in this dataset we do not have information on teacher gender beliefs to access whether they affect the pupil own beliefs about the reason of successful performance in math and reading.

Table D.I. Correlation between subject, gender, and reasons for performing well in exams

Dependent Variable: Reasons for performing well						
	Being good			**Exerting effort**		
	Math	Reading	Reading and Math	Math	Reading	Reading and Math
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.076***	-0.023***		0.139***	0.089***	
	(0.005)	(0.006)		(0.007)	(0.007)	
Test math	0.031***			0.010***		
	(0.002)			(0.002)		
Test reading		0.005***			0.049***	
		(0.002)			(0.002)	
Math			-0.026***			0.080***
			(0.009)			(0.010)
Math*Female			-0.062***			0.031**
			(0.011)			(0.013)
Individual FE	No	No	Yes	No	No	Yes
Obs.	16952	16979	34028	16952	16979	34028
R^2	0.042	0.001	0.531	0.025	0.038	0.605

Notes: The specific questions asked at the end of the standardized test in grade 6 for the cohort of students graduating from middle school in 2013-14 are: *(1) In a math exam, you correctly solved all exercises. Why does it happen? (2) You present to your classmates a text you have read. You do it in a clear and precise way and everybody follows it with interest. Why does it happen?* The potential answers are: *a)I was helped b)I was lucky c)It was easy d)I'm good e)I put in a lot of effort.* Robust Standard Errors clustered at class level are in parentheses. This information is collected together with the standardized test score in grade 6. "Female" refers to the gender of students. "Math" is a dummy which assumes value 1 if the dependent variable is related to own assessment in math and 0 if it is related to own assessment in reading. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

E Conceptual Framework

I develop a simple conceptual framework, similar to Dee (2014) and based on the stereotype threat idea developed in social psychology that can help interpreting the results. In this model, student beliefs about own ability in math are a function of own true unobserved ability a_i and teachers' gender stereotypes s_t, defined as follows: $\alpha_i = f^i(a_i, s_t)$. The impact of the bias on own self-perception is an individual specific function: students with higher vulnerability to the stereotype that "girls are not good at math" will be more negatively impacted by teacher stereotypes. This simple framework is flexible enough to capture heterogeneous perception of own ability among students with the same true unobserved ability a_i. For instance, a boy may perceive higher ability in math compared to a girl with the same unobserved talent for math a_i in the same class (i.e., exposed to the same teacher with stereotypes s_t). I assume that students' beliefs about own ability are a weakly decreasing function of teacher stereotypes, i.e., $\alpha_s \leq 0$. This is a testable assumption, which is supported by the empirical evidence available in Section 6 of the paper.

In a simple framework, students choose effort and individual's utility can be represented as

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

where u is differentiable and the sufficient conditions for a local interior maximum hold, k is the benefits function, which depends on ability (α_i) and effort (e_i), and the cost c is paid according with the level of effort e exerted by individual i. The component θ_i introduces an exogenous heterogeneity and it captures observable difference across individuals in the returns to performance. In this simple framework, I do not introduce parametric assumptions on the utility function.

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

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

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

Effort and perceived ability are often considered as complementary in the education production function (i.e. $k_{e\alpha} > 0$), so that a higher self-assessment of own capacities enhances the motivation to exert effort (Bénabou and Tirole, 2002). Hence, higher stereotypes will decrease the level of effort in equilibrium ($e_s^* \leq 0$). However, if the student increases effort as a reaction to a negative stance of the math teacher (i.e. $k_{e\alpha} < 0$), then the impact of stereotypes on effort is positive. As suggested also by Dee (2014), $e_s^* \geq 0$ is likely, for instance if individuals of the stigmatized group consider the stereotype strongly improper and react with an *"I'll show you are wrong"* attitude. In the context of gender stereotypes, it would imply that talented female students may increase the level of effort when they interact with teachers with stronger bias in order to disprove the negative belief.

In the empirical counterpart of this model, I observe improvements in achievement test scores (P) and not directly effort[2], but I assume for simplicity that the derivative of performance with respect to effort is positive $(P_e > 0)$ and I focus on the choice of the latter. Indeed, in the paper, I analyze whether improvements in achievement test scores are affected by teacher stereotypes. Assume two students, with the same gender, family background and math performance, are quasi-randomly assigned to two different teachers, respectively with stereotypes s_{t_i} and s_{t_j}, such that $s_{t_i} < s_{t_j}$. Then, if effort is complementary of students' perceived ability, the optimal level of effort (and therefore performance) of the student decreases with teacher stereotypes. However, if $k_{e\alpha} < 0$, then $e_s^* > 0$. This theoretical result may explain why girls at the top of the initial ability distribution have slightly higher, albeit indistinguishable from zero, improvements in math when exposed to teachers with stronger gender stereotypes.

E.1 Extension of the Conceptual Framework

I extend the simple conceptual framework to analyze the impact of teachers' gender stereotypes on effort of students, as mediated by both student perception of own ability (α_i) and teacher investment toward pupils, in the form of either time or encouragement, (β_{t_i}). The latter is an additional channel through which teacher bias (s_t) may impact students' performance and choices. I define β_{t_i} as an individual specific function of (s_t): $\beta_{t_i} = h^i(s_t)$. Furthermore, I assume that teachers with higher stereotypes are less supportive toward member of the stigmatized group, i.e., $\beta_s \leq 0$. Unfortunately, I do not observe data on gender specific investment or interaction in the classroom between professors and students, but the social psychology literature described in Section 6 provides evidence in support of this assumption.

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

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

$$u_i = \theta_i k(\alpha_i, e_i, \beta_{t_i}) - c(e_i) \tag{3}$$

where $\beta_{t_i} = h^i(s_t)$ and all other parameters and functions are defined as in equation (2).

The optimal level of effort with respect to teacher stereotypes is given by:

$$e_s^* = \frac{\theta(k_{e\alpha}\alpha_s + k_{e\beta}\beta_s)}{-(\theta k_{ee} - c_{ee})} \gtreqless 0 \tag{4}$$

In this extended framework, whether students increase or decrease their effort level when exposed to more biased teachers will depend on the complementarity and substitutability of effort with both own perceived ability $(k_{e\alpha})$ and teacher behaviour $k_{e\beta}$. If both are complementary, then the student will decrease the level of effort when exposed to a teacher with higher bias $(e_s^* < 0)$. If both are substitutes, we are in the case in which students work harder when exposed to more biased teachers to disprove the negative belief.

F Additional Outcome: Retention rate

In the Italian schooling system, at the end of each academic year, teachers decide whether the student is admitted to the following grade. This decision is based on the overall assessment of students, including both performance and behavior in class. The retention rate of males is higher compared to the one of females. For instance, among students in my sample, 7.7% of males and 4.7% of females are retained in at least one of the three years of middle school. In Table F.I, I check whether math teachers' bias has an impact on retention rate in the last year of middle school, when the probability of school failure is on average 0.2% for boys and 0.1% for girls. I find suggestive evidence that girls in classes assigned to math teachers with more gender stereotypes are slightly more likely to fail (Panel A), while there is no significant effect for literature teachers (Panel B). However, these effects are driven by very few cases.

Table F.I. Estimation of the effect of teachers' gender stereotypes on grade retention in grade 8 - class FE regression

	(1)	(2)	(3)	(4)	(5)
Panel A- Dependent Variable: grade retention at the end of grade 8					
Female	-0.001*	-0.000	-0.000	-0.004**	-0.007*
	(0.000)	(0.000)	(0.000)	(0.002)	(0.004)
Fem*Math Teacher Stereotypes		0.001*	0.001*	0.001*	0.001*
		(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.002***	0.002***	0.002***	0.004***	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Obs.	24840	24840	24840	24840	24840
R^2	0.076	0.076	0.077	0.077	0.078
Panel B- Dependent Variable: grade retention at the end of grade 8					
Female	-0.001**	-0.001*	-0.001*	-0.005***	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
Fem*Lit Teacher Stereotypes		-0.000	-0.000	-0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.001)
Constant	0.002***	0.002***	0.003***	0.005***	0.005***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Obs.	23804	23804	23804	23804	23804
R^2	0.073	0.073	0.074	0.075	0.075
Class FE	Yes	Yes	Yes	Yes	Yes
Student Controls	No	No	Yes	Yes	Yes
Student Controls*Fem	No	No	No	Yes	Yes
Teacher Controls*Fem	No	No	No	No	Yes

Notes: This table reports OLS estimates of equation 1, where the dependent variable is math or reading standardized test score in grade 8 in Panel A and B, respectively, and the dependent variables refer to math teachers in Panel A and to literature teachers in Panel B. The unit of observation is student i in class c taught by teacher t in grade 8 of school s. Standard errors (in parentheses) are robust and clustered at the teacher level. The variable "Fem" indicates the gender of the student. Individual controls include education of the mother, occupation of the father, immigrant dummy, generation of immigration, and their interactions with the gender of the student. Teacher controls include the interaction between student gender and teacher gender, place of birth, age, children and daughters, advanced STEM degree (physics, math, engineering), leader of school math Olympics, degree cum laude, refresher courses, type of contract, and education of the teacher's mother. *, **, and *** indicate significance at the 10%, 5% and 1% percent level respectively.

References

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

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

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

Greenwald, A. G., B. A. Nosek, and M. R. Banaji (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology 85*(2), 197.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology 141*(2), 165–173.

Riegle-Crumb, C. and M. Humphries (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society 26*(2), 290–322.

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