Appendix for Online Publication

Additional Tables and Figures

Table A. 1: Probability of Teaching the Same Class Long-Term

Teacher G-Styping	-0.000	0.007	-0.009	0.013
	(0.04)	(0.04)	(0.05)	(0.05)
Male		-0.086	-0.041	-0.052
		(0.10)	(0.10)	(0.10)
Years of Experience		0.004	0.001	-0.001
		(0.01)	(0.01)	(0.01)
Education Degree		-0.009	-0.006	-0.046
		(0.17)	(0.16)	(0.19)
Linguistics		0.391*	0.565***	0.528***
		(0.22)	(0.15)	(0.15)
Natural Sciences		-0.115	-0.088	-0.106
		(0.18)	(0.17)	(0.14)
Social Sciences		-0.190	-0.158	-0.155
		(0.13)	(0.13)	(0.13)
Growth Mindset			-0.026	-0.021
			(0.02)	(0.02)
Extrinsic Motivator			-0.017	-0.010
			(0.02)	(0.02)
Modern Approach			0.032***	0.028**
			(0.01)	(0.01)
Warm Approach			0.001	0.004
			(0.02)	(0.02)
Number of Extra_C Programs				0.004
				(0.01)
Number of Volunteer Activities				0.021
				(0.01)
N	145	145	145	145
Pseudo-R-squared	0.00	0.05	0.09	0.11
Significance of model test	1.00	0.28	0.04	0.02

Reported estimates are average marginal effects from logit regressions. The dependent variable "Long-term teaching" is defined as a binary variable, which takes the value 1 if the teacher has been teaching the same class for 4 years in grade 4 and 3 years in grade 3. Heteroskedasticity-robust standard errors are in parentheses. The last row "Significance of model test" gives the p-value for joint significance of all covariates used in the corresponding specification.

Table A. 2: Teacher Gender Role Beliefs and Student Scores - Details of Table 3 in the main text

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.000	-0.055	0.054	-0.094*
	(0.06)	(0.05)	(0.04)	(0.05)
Medium Term	0.022	0.058	0.026	0.033
	(0.07)	(0.06)	(0.07)	(0.08)
Long-Term	0.117	0.193***	0.015	0.006
	(0.08)	(0.07)	(0.08)	(0.07)
Medium Term*Teacher G-Styping	-0.120*	0.001	-0.110**	0.020
	(0.06)	(0.06)	(0.05)	(0.07)
Long Term*Teacher G-Styping	-0.211**	-0.016	-0.162**	-0.026
	(0.08)	(0.07)	(0.07)	(0.07)
Student Characteristics:				
Age(months)	0.005*	0.004	-0.001	0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Raven Score	0.350***	0.225***	0.298***	0.257***
	(0.03)	(0.02)	(0.02)	(0.02)
Teacher's assesment: well-behaved	0.110***	0.137***	0.151***	0.127***
	(0.03)	(0.02)	(0.03)	(0.02)
Academic Self-confidence	0.129***	0.155***	0.079***	0.115***
	(0.02)	(0.02)	(0.02)	(0.02)
Student G-Styping	-0.138***	-0.116***	-0.119***	-0.124***
	(0.02)	(0.02)	(0.02)	(0.02)
Student GMS	0.018	0.037*	0.085***	0.068***
	(0.02)	(0.02)	(0.02)	(0.02)
Family Characteristics:				
Middle SES	0.108**	0.127**	0.068	0.140**
	(0.05)	(0.05)	(0.05)	(0.05)
High SES	0.186***	0.141**	0.158***	0.197***
	(0.05)	(0.06)	(0.06)	(0.07)
Working Mother	0.017	0.038	0.087*	-0.029
	(0.05)	(0.04)	(0.05)	(0.05)
Computer at Home	0.037	0.152***	0.004	0.105**
	(0.04)	(0.05)	(0.05)	(0.04)
G-Styping at Home	0.034*	0.022	0.024	-0.008
	(0.02)	(0.02)	(0.02)	(0.02)
Teacher Characteristics:				
Male Teacher	-0.033	0.120**	0.081	0.046
	(0.07)	(0.06)	(0.06)	(0.07)
Teacher Qual - 2 Year College	-0.252*	0.224**	0.111	0.116
	(0.14)	(0.10)	(0.11)	(0.13)
Teacher Qual - Grad S	-0.018	-0.014	0.050	0.010
	(0.10)	(0.10)	(0.10)	(0.07)
Years of Teaching	0.006	-0.002	-0.010**	-0.008*
	(0.00)	(0.00)	(0.00)	(0.00)
Linguistics	0.090	0.201*	0.180	-0.039
	(0.13)	(0.12)	(0.11)	(0.15)
Sciences	-0.100	0.088	0.159*	0.064
	(0.09)	(0.08)	(0.08)	(0.09)
Social Sciences	-0.181	0.071	-0.131	-0.014
	(0.13)	(0.11)	(0.13)	(0.13)
Other	-0.160	-0.011	0.045	-0.002
	(0.12)	(0.12)	(0.10)	(0.09)
Teacher Styles:				
GMS	-0.041***	-0.035***	0.013	-0.021
	(0.01)	(0.01)	(0.01)	(0.01)
Extrinsic Motivation	0.005	-0.018	0.008	-0.011
	(0.02)	(0.01)	(0.01)	(0.01)
Modern Approach	-0.006	0.011	0.010	0.001
	(0.01)	(0.01)	(0.01)	(0.01)
Teacher Warmth	-0.021	-0.033***	-0.003	-0.010
	(0.01)	(0.01)	(0.01)	(0.01)
Teacher Effort:				
Occupational Trainings	-0.000	-0.000	-0.002	-0.005
	(0.00)	(0.00)	(0.00)	(0.00)
Extra-curricular	0.003	0.015***	-0.004	0.012**
	(0.01)	(0.01)	(0.01)	(0.00)
School Fixed Effects	✓	✓	✓	✓
N	1870	1943	1873	1946
R-Squared	0.32	0.34	0.26	0.26

Table A. 3: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores, Excluding Very Progressive Teachers

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.027	-0.028	0.069	-0.102*
	(0.08)	(0.06)	(0.05)	(0.06)
2-3 Year Exposure	0.006	0.048	0.009	-0.005
	(0.07)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.137	0.200***	-0.029	-0.021
	(0.09)	(0.07)	(0.07)	(0.08)
2-3 Year Exposure*Teacher G-Styping	-0.135*	-0.021	-0.124**	0.061
	(0.07)	(0.07)	(0.06)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.231**	-0.007	-0.102	0.044
	(0.10)	(0.09)	(0.09)	(0.08)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.248	0.861	0.782	0.803
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.061	0.763	0.042	0.404
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.028	0.937	0.257	0.556
P-value: 1 Year E*G-Sty[Girls=Boys]	0.467		0.007	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.219		0.781	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.063		0.740	
N	1729	1798	1732	1801
R-Squared	0.33	0.34	0.25	0.26

Dependent variables are standardized test scores. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level. Teachers who scored lower than the 10th percentile (very progressive) are excluded.

Table A. 4: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores: Teachers with Less than 20 Years of Service

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	-0.010	-0.029	0.065	-0.090*
	(0.06)	(0.05)	(0.04)	(0.05)
2-3 Year Exposure	-0.016	0.027	-0.011	0.015
	(0.07)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.089	0.213***	-0.015	-0.013
	(0.09)	(0.07)	(0.08)	(0.07)
2-3 Year Exposure*Teacher G-Styping	-0.122*	-0.005	-0.126**	0.031
	(0.07)	(0.06)	(0.05)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.222**	-0.031	-0.177**	-0.044
	(0.09)	(0.07)	(0.08)	(0.07)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.202	0.686	0.454	0.224
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.063	0.926	0.021	0.642
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.011	0.678	0.024	0.555
P-value: 1 Year E*G-Sty[Girls=Boys]	0.758		0.003	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.038		0.953	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.017		0.728	
N	1695	1755	1698	1758
R-Squared	0.32	0.33	0.25	0.26

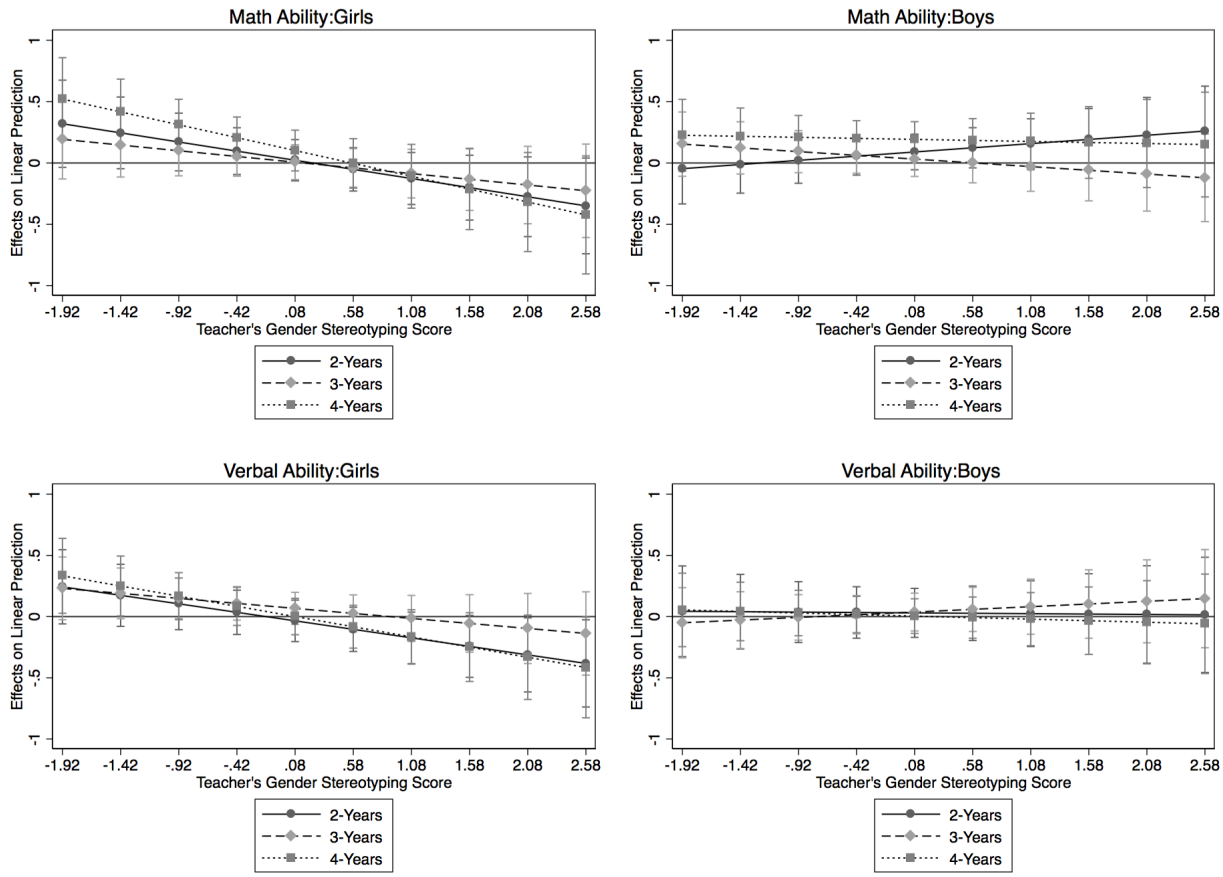
Dependent variables are standardized test scores. Estimated coefficients are obtained by constraining the sample to teachers who have less than 20 years of service. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student- reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher Characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Table A. 5: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores: Excluding Teachers Who Believe Boys are Better at Math

	Math Score		Verbal Score	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.008	-0.057	0.063	-0.090
	(0.07)	(0.05)	(0.05)	(0.06)
2-3 Year Exposure	0.006	0.062	0.036	0.014
	(0.08)	(0.07)	(0.07)	(0.08)
4 Year Exposure	0.098	0.210***	0.017	0.004
	(0.09)	(0.08)	(0.08)	(0.08)
2-3 Year Exposure*Teacher G-Styping	-0.113*	0.008	-0.127**	0.040
	(0.07)	(0.06)	(0.06)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.208**	-0.009	-0.171**	-0.020
	(0.08)	(0.08)	(0.08)	(0.08)
School Fixed Effect	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.215	0.781	0.496	0.329
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.097	0.899	0.029	0.579
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.015	0.904	0.031	0.802
P-value: 1 Year E*G-Sty[Girls=Boys]	0.324		0.007	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.253		0.751	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.084		0.992	
N	1772	1836	1775	1839
R-Squared	0.32	0.34	0.26	0.27

Dependent variables are standardized test scores. Estimated coefficients are obtained by dropping the teachers who reported that, in their teaching experience, they observed that boys were better at math than girls. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mindset, teacher-reported behavior score. Family characteristics: student-reported gender roles at home, mother's employment status, teacher-reported socioeconomic status categories. Teacher characteristics: teacher gender, tenure, education, experience, branch of study (social sciences, linguistics, humanities, science and teaching). Teaching styles: Scores constructed for warm vs. distanced, extrinsic vs. intrinsic motivator, traditional vs. modern and growth vs. fixed mindset. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level.

Figure A. 1: Effects on Test Scores: Linear Predictions



Figures present linear predictions (and 95% confidence bands) obtained from estimating the specification presented in Table 3 in the main text, where teacher stereotyping score is interacted with exposure categories (1,2,3 and 4), rather than 1, 2-3 and 4.

Questions Used for Constructing Gender Role Beliefs–Students

4-point item scale: completely agree, agree, disagree, completely disagree

1. It is the father's responsibility to earn a living in a family, and it is the mother's responsibility to take care of the children.
2. Being a nurse is not a suitable profession for a man.
3. Men cannot sew well even if they try hard to learn it.
4. Women cannot play football well even if they try hard to learn it.
5. Men are better at math than women.
6. Being a space scientist/astronaut is not a suitable profession for a woman.
7. It is more natural for girls to help with housework than boys.

Questions Used for Constructing Gender Role Beliefs

4-point item scale: completely agree, agree, disagree, completely disagree

1. It is the father's responsibility to earn a living in a family, and it is the mother's responsibility to take care of the children.
2. Being a nurse is not a suitable profession for a man.
3. Men cannot sew well even if they try hard to learn it.
4. Women cannot play football well even if they try hard to learn it.
5. Men are better at math than women.
6. Men generally understand money-related issues better than women.
7. It is much more important for boys to go to university than girls.
8. It is more natural for girls to help with housework than boys.
9. Men have better judgment compared to women, hence they are better leaders.

Questions Used for Constructing Teaching Styles

4-point item scale: completely agree, agree, disagree, completely disagree

Growth vs. Fixed Mindset

1. Intelligence is a fixed trait. One cannot change how smart he/she is.

2. People can improve their intelligence regardless of their innate level.

3. Only very few people can excel in arts, music and sports, as innate ability is required to be successful.

4. Working hard does not make you successful in a task unless you are talented.

5. If a student works hard enough, he/she can be the best in the class.

Extrinsic vs. Intrinsic Motivator

6. Punishment is necessary to attain a disciplined and ordered classroom environment.

7. I often reward students (applauding, giving stars etc.) to elicit the outcomes and behaviors that I aim for.

8. I often punish students (grounding them on the breaks, making them sit alone etc.) to elicit the outcomes and behaviors that I aim for.

9. Rewarding behaviors or outcomes with material incentives (giving them stars and stickers etc.) prevents students from developing intrinsic motivation.

Modern Approach vs. Traditional Approach in Teaching

10. A noisy classroom is not a problem as long as students are busy with learning.

11. It is important to let students express their ideas regardless of how wrong and absurd they are.

12. I do not like to fall behind on the syllabus due to students' problems and questions or any other reason.

13. It is more efficient to teach students the correct answers directly rather than asking them questions and spending time on their potentially wrong answers.

14. Students should be entitled to choose what activities we do in the class.

15. When a student asks a question about a subject he/she is curious about, I only answer it if it is related to the subject I am covering at that moment. If it is irrelevant, I leave it to a later time not to disrupt the class flow.

Warm vs. Distanced

16. Teachers should keep their distance and be the authority in their relationship with the students, as this is beneficial for the students' development.

17. My educational standards and expectations from students can be described as strict and prescriptive.

18. Inculcating a strict discipline and ability to obey in students during elementary school, despite being difficult, is very beneficial for them further in their lives.

19. Having a warm teacher-student relationship and a classroom environment where students feel comfortable is more important for effective learning than a respect-based teacher-student relationship and a quiet classroom.