GENDER STEREOTYPES IN THE CLASSROOM AND EFFECTS ON ACHIEVEMENT

Sule Alan,

University of Essex

Seda Ertac,

Koc University

and

Ipek Mumcu

University of Essex

Abstract: We study the effect of elementary school teachers' beliefs about gender roles on student achievement. We exploit a natural experiment where teachers are prevented from self-selecting into schools, and conditional on school, students are allocated to teachers randomly. We show that girls who are taught for longer than a year by teachers with traditional gender views have lower performance in objective math and verbal tests, and this effect is amplified with longer exposure to the same teacher. We find no effect on boys. We show that the effect is partly mediated by teachers transmitting traditional beliefs to girls.

JEL Categories: I21, J16, J24.

Acknowledgement: We would like to thank ESRC (through MiSoc) and ING Bank (through its Corporate Social Responsibility programme) for support. Seda Ertac acknowledges support from the Turkish Academy of Sciences (TUBA, GEBIP program) and the Science Academy (BAGEP program). We thank Thomas Crossley, Imran Rasul, Barbara Petrongolo, Greg Fischer and participants at the CEPR/IGC Conference in Labor Markets in Developing Countries at the London School of Economics, participants at the AEA/ASSA

2017 Chicago Meeting, and seminar participants at the University of Bristol for helpful comments. We are grateful to numerous research assistants who helped us collect the data. All errors are our own.

1 Introduction

Stereotypes about gender are pervasive in most societies. These views tend to rigidly define the innate capabilities and attitudes of each sex, and social roles that are deemed appropriate for men and women. To the extent that they influence the actual choices and outcomes of individuals, such beliefs may in large part contribute to gender-achievement gaps as well as the underrepresentation of women in top executive positions, STEM careers, and in leadership. As ample evidence suggests, such gender inequality, factually confirming and perpetuating traditional gender role beliefs, can be quite persistent (Bertrand and Hallock (2001), Blau, Ferber and Winkler (2002), Fortin (2005), Bertrand (2011)).

The formation of gender role beliefs and conforming behaviors and attitudes likely begins very early in childhood, within the family, as families have the earliest, most direct impact on children's beliefs and preferences (Bisin and Verdier (2001)).[1] Once a child starts school, factors that contribute to the formation of beliefs and attitudes become broader and more complex. In addition to their families, children now interact with their peers in a more structured environment and perhaps more importantly, with another adult, the teacher. Teachers' views toward gender roles may affect students' attitudes, behaviors and outcomes both directly and indirectly. First, a teacher's beliefs may influence students' achievement outcomes by influencing students' own beliefs: the teacher may express his/her views about

[1]It has been documented that transmission of gender attitudes from mothers affects daughters' as well as daughter-in-laws' labor force participation and human capital (Farre and Vella (2013), Johnston et al. (2014)). Olivetti et al. (2016) find that women's work hours are positively affected both by the work behavior of their own mother and their peers' mothers.

gender-appropriate roles in the classroom, and because he/she is a significant authority figure, students may adopt and internalize what their teacher says. These beliefs may in turn influence girls' academic aspirations, their interest in male-stereotyped topics such as math, and their motivation to study for as well as the level of stereotype threat and anxiety they may experience in subjects in the male domain (Spencer et al. (1999)).

A more direct mechanism is the teacher interacting differently with girls and boys. A teacher with strongly traditional gender role beliefs may think that acquiring academic skills is not as important for girls, since they are unlikely to put them into practice later in life. Such a teacher may reflect these beliefs in actual classroom practices, by giving different types of feedback to girls and boys, selectively answering/dismissing questions, or focusing on boys when teaching (Sadker and Sadker (2010)). Biases on the part of teachers can also manifest through discrimination in grading (either against or in favor of girls), and this can affect student achievement and choices (Lavy and Sand (2015), Terrier (2015), Lavy (2008)). In addition to directly influencing learning, such teaching practices on the part of biased teachers can affect long term outcomes by affecting the development of girls' non-cognitive skills as well.[2] A very progressive teacher, on the other hand, may exert extra effort to engage students in subjects that are typically considered in the domain of the opposite sex and try to break stereotypical attitudes in the classroom.

In this paper, we study the effect of teachers' beliefs about gender roles on their students' achievement outcomes, using rich data from a large-scale field study involving approximately 4000 3rd and 4th grade students and their 145 teachers. In order to identify these effects, we exploit the unique institutional features of our study site, Turkey. The educational system in Turkey provides us with a natural experiment with three main components: First, stratified by gender and pre-school education, state elementary school students are allocated to their teacher in first grade randomly. Second, teachers are appointed to schools centrally by the

[2]It is well-known that non-cognitive skills in childhood are predictive of many important outcomes over the life cycle (Almlund et al. (2011)).

Ministry of Education based on the need for teachers and as such, they are prevented from self-selecting into catchment areas and schools before acquiring a considerable number of years of service. Finally, the general practice is such that students have the same teacher for the entire elementary school period, from grade one to four, and this is disrupted mainly by teacher rotations and, to a lesser extent, family relocations. This disruption provides us with variation in the number of years a student in a school is taught by the same teacher, allowing us to identify the mediating effect of length of exposure to a teacher with particular gender role beliefs. We provide details on these institutional features in Section 2.

The data reported in this paper were collected by the authors as part of a large field study, with the specific goal of exploring the role of the elementary school teacher in shaping children's beliefs and affecting their achievement outcomes. As such, the dataset includes a rich set of variables on student, family and teacher characteristics, which were collected by physically visiting the classrooms several times. Having access to a rich set of teacher quality indicators was our primary motivation in our data collection effort. This is because teachers' gender role beliefs are likely correlated with teaching quality, rendering the identification of the effect of these beliefs on achievement outcomes difficult.[3] A particular strength of our data is that detailed information on teachers with respect to their daily classroom practices, teaching styles, and their pedagogical approach to teaching as well as indicators of personal effort are collected through surveys.

We find that teachers' gender role beliefs have quite different effects on girls and boys. Girls taught by teachers with traditional views about gender roles for more than one year have lower performance in objective math and verbal tests, and this effect is amplified in size with longer exposure to the same teacher. If the teacher has been teaching the student

[3]There is a large literature in economics that studies the effect of teacher quality on educational attainment (e.g. Rockoff (2004), Rivkin et al. (2005), Aaronson et al. (2007), Harris and Sass (2011), Hanushek (2011), Chetty et al. (2014)). See also Hanushek and Rivkin (2006), Schwerdt and Wuppermann (2011)).

for two to three years, a one standard deviation increase in teacher stereotypes leads to a 0.12 and 0.06 standard deviation decrease in math and verbal test scores, respectively. This negative effect becomes 0.21 for math and 0.11 for verbal test scores if the student is taught by the same teacher for the entire duration of elementary school (four years). We find no statistically significant effect of teachers' gender role views on boys' test scores.

We then explore various channels through which teachers' beliefs may affect girls' test scores. Our statistical mediation analysis shows that about 17% of the effect of a traditional teacher on girls' math test scores is coming from girls' gender roles beliefs being influenced by their teacher's gender role beliefs. Other potential mechanisms notwithstanding, our results suggest that teachers' influence on girls' beliefs on gender roles may be an important indirect channel. To the extent that these beliefs predict important real life outcomes such as choice of study major and occupation, we conjecture that the importance of this channel extends well beyond test scores.

The role of teacher gender has been an important focus in the education literature, and it has been shown that having a female teacher may affect outcomes such as math performance, STEM grades and graduation rates on the part of female students (Bettinger and Long (2005), Dee (2007), Hoffmann and Oreopoulos (2009), Carrell et al. (2010), Antecol et al. (2014)). It has also been shown that the student-teacher gender (mis)match can influence a teacher's perceptions of the student (Dee (2005)). The effect of teachers' beliefs and attitudes with respect to gender roles, however, has received less attention. As mentioned above, a recent set of papers document effects of gender biases as reflected in discrimination in grading on student achievement and choices, with differing results (Lavy and Sand (2015), Terrier (2015)). The former paper finds (in Israel) that boys are over-assessed, with negative effects on girls' achievement and future math course choices, while the latter finds (in France) that girls are favored in grading in math, and this increases girls' propensity for choosing a science track in high school. In addition, Lavy (2008) documents an anti-male bias in grading and Robinson and Lubienski (2011) also find that teachers rate girls more favorably

than cognitive scores would suggest. The current paper differs from these studies in that we measure teachers' gender role beliefs directly rather than using grading biases, use variation in the duration of exposure to the teacher, and control for teaching quality and styles, which can be correlated with both teachers' gender attitudes and students' achievement. The paper contributes to the literature on teacher effects on achievement by showing that teachers' beliefs and attitudes are important in determining achievement outcomes and gender gaps in those outcomes, as well as in shaping the beliefs and attitudes of students. Our data, comprising teacher and student characteristics which are typically not available, allow us to construct a continuous measure of gender stereotypes to facilitate non-parametric as well as parametric identification. The unique educational setting allows us to estimate the mediating effect of the length of contact with a particular type of teacher. Our results highlight that the classroom environment, in particular the type of teacher, is an important part of a child's social environment and already starts influencing children's performance and beliefs at the elementary school level. The results broadly suggest that gender-equal classroom practices, implemented early on by teachers with progressive views, could prevent gender gaps in achievement that likely cause multiplicative effects on academic persistence, occupational selection and labor market outcomes later in life.

2 Background

The Turkish 12-year compulsory education is based on a two-tier system, where both public and private schools are under the oversight of the National Ministry of Education. As Turkey has moved from low income to middle income status over the last 15 years, the majority of the middle- and upper-class parents prefer to send their children to private schools. Our study sample covers 3rd and 4th grade students in state-run (public) elementary schools in particularly needy areas of Istanbul. As such, it primarily represents Turkey's lower socioeconomic segment, with limited variation with respect to socioeconomic status.

In studying the impact of teachers' beliefs on the actual outcomes of students, one faces a

fundamental selection issue, that is, students in a given school may be allocated to teachers in a non-random manner. This happens, for example, when a particular type of parent selects a particular type of teacher, a teacher known to be better or appearing to have similar beliefs and attitudes as the parent. If gender role beliefs somehow proxy unobserved teacher quality, for example if more progressive teachers are also more likely to use modern teaching methods or adopt a more constructive approach, or they are simply more intrinsically motivated and care more about their students' achievement, such selection compromises identification. Our setting circumvents this selection issue.

After the registration of all first-graders (school-starters) in a given academic year, school administrators randomly allocate the students to teachers through publicly held draws in the presence of parents. Classroom sizes are not allowed to exceed 50, although a maximum of 30 is typically preferred. Draws are stratified based on gender and pre-school attendance to ensure balance in gender and school-preparedness in each classroom. Therefore, contrary to the private school system, there is no room for parents to choose their child's teacher in the state system. Of course, a parent may decide to send their child to a school that is not in the catchment area; however, acceptance of the student to a non-catchment area school is subject to the capacity of that school and priority is given to catchment area residents. Sending the child to a school that requires transportation is costly and relocations for educational purposes are extremely rare in this socioeconomic group. This, along with centrally managed teacher appointments, ensures that exposure to the same teacher is largely independent of teacher and student quality. Once students are allocated to classrooms in grade 1, re-mixing in later grades is extremely rare, which means that students remain with the same classmates until they graduate from elementary school, unless the family moves.

Despite the random allocation of students to teachers, if our gender stereotype construct is correlated with some unmeasured aspect of teacher quality, it would still be difficult to interpret our results as the causal effect of teacher's beliefs on student achievement. To isolate the effect of beliefs as much as possible, we collected very detailed information on

teachers. In addition to demographic characteristics, these include the teacher's teaching philosophy, pedagogical approaches, classroom practices and indicators of effort and care for student achievement. We explain how we construct summary measures based on this information in Section 3.2.

The final issue to account for in studying the effect of beliefs on actual outcomes is the fact that such effects, if they exist, may take a long time to surface. It is plausible that the longer the exposure to the same teacher, the larger and more persistent the effects may be. In many countries, elementary school students are taught by a different teacher each year, making it difficult to detect teacher effects. However, this is not the case in our study site. Except for involuntary rotations, re-appointments and retirement, a teacher teaches the students allocated to him/her from grade 1 to grade 4, after which those students move on to middle school. Because of the strictly centralized allocation of teachers and subsequent re-appointments and rotations (explained below), we have substantial exogenous variation in the length of time a given student has spent with the same teacher, which gives us a unique opportunity to study the role of the length of exposure in moderating impacts.

Allocation of Teachers to Schools

Although we exploit only the within-school variation to estimate the effects, it is important to provide a brief account of the way teachers are allocated, rotated and re-appointed (centrally) in our study site. This is because the specific features of this system will provide support for our exogeneity assumption with respect to the time spent with the same teacher, which allows us to identify the mediating effect of exposure. After completing the degree requirement, the current practice in the public system is that all teacher candidates take a nationwide civil servant examination and those above a cutoff score are placed in a pool to be appointed to a public school in need.[4] A new teacher has typically no say in which city, let alone district

[4]Private schools, despite being subject to the curricular requirements of the Ministry of Education, enjoy autonomy in implementing their own teacher selection process, and are not

or school she will be appointed to. It is generally very difficult to be appointed to one's preferred city before 5 to 10 years, except for pure luck. In 2015, among over 300,000 new teachers, only 40,000 were appointed. The situation leaves no bargaining power to teachers as every year an increasing number of teachers remain unappointed, waiting for the next round of appointments.

Once appointed, teachers begin to collect service points that are assigned to their school. Each school has a score assigned to it by the Ministry, with schools in deprived and dangerous areas having higher scores than those located in well-off cities, districts within cities and catchment areas within districts. A teacher mechanically earns the points assigned to her school, for every year she teaches. There is no other way for a teacher to accumulate service points other than by simply teaching. These points are very important for teachers, as they determine their chances of being re-appointed to the city of their choice, or the district of their choice if they are already in a city they like.

After his/her appointment, a teacher can be re-appointed to another school (generally within the same city) if i) there appears to be an excess supply of teachers at her current school and she has the lowest service points among her colleagues (involuntary rotation), ii) her re-appointment request is honored.[5] A classroom may lose its teacher because of retirement and resignations but the most common reason is involuntary rotation due to

subject to the scrutiny of the Ministry in this regard.

[5]Teachers cannot ask to be re-appointed before completing at least 4 years (over 6 years in actual practice) of service in their current school. Requests to be re-appointed are honored if i) there is a school in need in the preferred district and ii) the teacher has higher service points than her competitors who have the same location preference. As working in high-SES catchment areas is more desirable for most teachers, there tends to be a high teacher turnover in low-SES district schools such as the ones that comprise our sample. For an Istanbul teacher, even with a long tenure in the profession, it is extremely hard to be appointed to the generally desired (high-SES) districts.

excess supply and re-appointment to another school based on teacher request. Note that when a teacher is re-appointed to a new school, she is allocated to a classroom which is in need of a teacher. As this classroom can be of any grade, such moves contribute to the variation we observe in the length of exposure to a given teacher from the point of view of the student.

While teachers who want to move (as our schools are in relatively remote and deprived areas, most in our sample say they would like to, once they accumulate sufficient points) do so mainly to work in the district of their choice, the centralized system makes it very difficult for them to self-select into catchment areas and schools conditional on district. Such self-selection becomes possible only for a teacher with very high service points, usually having taught beyond 25 years or more than the usual amount of time working in high-point areas such as eastern Turkey. While we base our identification strategy (conservatively) on within-school variation through the use of school fixed effects, it is important to re-iterate that teacher sorting within district based on any metric other than service years, which we control for in our regressions, is largely ruled out in this system. In our results section, we show that teachers who have been teaching a class for a longer and shorter time are largely similar in terms of the rich observables we have such as demographics, qualifications and teaching styles.

3 Data

Our data were collected as part of a large-scale field project, which has been underway since 2013. The project aims to study the behaviors, attitudes and outcomes of students in conjunction with the behaviors and attitudes of teachers. All student data were collected by the authors of the paper by physically visiting all classrooms multiple times.[6] We took great care to ensure that the teacher was not present when the students worked on the tests

[6]The project has local IRB approval as well as official state approval.

and filled in the questionnaires.

Data were collected using a rich battery of tools, which includes surveys, a fluid IQ test and official grade records as well as objective mathematics and verbal tests that we prepared and conducted in the classroom. This endeavor required visiting each classroom multiple times to minimize disruption to daily teaching activities. Because there tends to be about a 20% non-attendance on each day due to sickness or other valid excuses, we do have some missing data on students. Our analysis is based on the teachers and students for whom we have complete information on key variables, forming a dataset with 31 schools, 145 teachers and approximately 4000 students. Our typical teacher is female, university educated, and has accumulated about 15 years of service as a teacher. Only about 25% of our teacher sample is male, as teaching in elementary school is still predominantly a female profession in Turkey. A little over 70% of our teachers have majored in a program called "class teaching", which is a 4-year university degree in elementary school teaching.

Our typical student in grade 3 (4) is 9 (10) years of age, and on average 70% of all third-graders have been taught by the same teacher for 2-3 years and 30% for one year. The respective percentages for the fourth grade is 55% for more than three years, 24% for two-to-three and 21% for one year.

3.1 Student and Family Characteristics

To account for the role of student and family characteristics in determining academic achievement, we collect rich measures of behaviors, attitudes and beliefs, as well as demographic information, information about the home environment, socioeconomic status and family background. For this, we use survey data from the students themselves as well as from their teachers. In particular, teachers fill out an extensive survey for each individual student, which includes questions regarding the attitudes and behaviors of the student within the classroom, the teacher's assessment of the student's attitudes, traits and performance, and his/her assessment of the student's family characteristics such as socioeconomic background.

Student surveys also include questions regarding the student's home environment to better capture the socioeconomic status as well as the behaviors and attitudes of the parents.[7]

Our main outcome measure consists of standardized math and verbal (Turkish) tests, which we implement in each classroom in the absence of the teacher. These tests were prepared (and extensively piloted) by the authors of the paper based on the national curricula. An independent set of teachers were consulted to tailor the questions to each grade (3 and 4). Students' cognitive ability is measured via Raven's Progressive Matrices. We also have access to students' official math, verbal and behavior grades, all given by their own teachers.

3.2 Teacher Characteristics

The primary purpose of the paper is to show the effect of the teacher's gender role beliefs on students' achievement outcomes. However, we acknowledge that these beliefs are likely to be correlated with certain underlying teacher characteristics that are likely instrumental for student achievement. For example, without adequately controlling for teacher quality, even in the absence of selection, it is difficult to give the association between beliefs and achievement outcomes causal interpretation due to the plausible correlation between gender role beliefs and quality. While there is consensus that teacher quality matters a lot for achievement over and above student characteristics (cognitive and non-cognitive skills) and family background, it has proven to be very difficult to measure it.[8] This is possibly because teacher quality is multidimensional, often involving unobservables such as teaching styles, effort and care. Acknowledging this difficulty, we collect two sets of additional information from our teachers, with the hope of better capturing the often unobserved components of

[7]We did not attempt to collect these information directly from parents, as our previous experience shows that the response rate of parents is very low and their answers to the surveys questions are usually not reliable. Instead, we rely on the child and the teacher for this information.

[8]See Carrell and West (2010).

teaching quality.

First, in addition to their education, experience and study majors, we collect a set of variables that relate to the teaching styles and pedagogical approach of our teachers. Teachers' styles of teaching the class material and interacting with their students, as well as their expectations from the students, are likely to be important factors in student outcomes (Domino (1971), Schwerdt and Wuppermann (2011), Bietenbeck (2014), Hidalgo-Cabrillana and Mayan (2015)). Using item-set questions directed to teachers, we construct four distinct teaching style variables. We call these modern vs. traditional, growth- vs. fixed-mindset, warm vs. distanced, and extrinsic vs. intrinsic motivator. A traditional teaching style is reflected in the teacher dictating to the students what to do in class, and following a rigid structure to each class that is determined by the teacher. What we call a modern approach to teaching, on the other hand, involves the students more in the learning process and aims to induce the children to think critically.[9] Having a growth mindset (Dweck (2006)) is the belief that abilities are malleable and success can be achieved provided that sufficient effort is exerted, regardless of innate characteristics. Such a mindset has been found in the literature to predict academic achievement (Blackwell et al. (2007), Alan, Boneva and Ertac (2016)). From the perspective of the teacher, we measure growth mindset through questions about the relative importance of innate ability vs. sustained effort for success, e.g. whether or not the teacher agrees that any student could become the best in the class if he/she works hard enough. The warm vs. distanced construct gets at how authoritarian the teacher is in his/her interactions with the students, and how important it is for him/her to establish a close and warm relationship with them. Finally, extrinsic motivator refers to the use of extrinsic rewards in motivating students (such as stickers, small gifts and applause for good performance) and punishment for inducing desired behavior. The full inventory we use to

[9]Estimating the effects of traditional vs. modern teaching practices on achievement has been an active research topic in the economics of education (e.g. Bietenbeck (2014)).

construct each style score is given in the Online Appendix.[10]

In addition to teaching styles, a crucial variable to control for is teaching effort or how much teachers care for students' achievement. However, the motivation and effort level of teachers are difficult to observe. The educational system we study, where there are no extrinsic incentives for teachers to maintain a high level of teaching, makes intrinsic motivation somewhat easier to measure, since any extra-curricular activity done by teachers reflects voluntary effort.[11] We therefore collect information on teachers' extra-curricular activities that focus on teaching improvement and student achievement through our survey. We believe this is informative of the teacher's (typically unobserved) care and effort in our setting. This is because, as mentioned before, teachers collect service points passively, by teaching only. No other activity or certificate or diploma will matter in collecting service points, which is required for re-appointments, salary increases and retirement benefits. Having said that, there are many certificate and diploma programs as well as conferences and social projects that aim to inform teachers about best classroom practices based on new evidence, with the goal of improving student achievement. Teachers who participate in these programs do so in a voluntary manner, paying participation fees (if any) themselves and sacrificing time during their evenings and weekends. Similarly, teachers do not gain anything other than professional satisfaction by organizing educational class trips, which often cost them money and require considerable effort, mainly because of the lack of parental interest in the socioeconomic segment we cover. We take the reported frequency of these volunteer activities as measures of teacher effort.

[10]Some of these questions were adapted from the "Teaching and Learning International Survey" (TALIS) questionnaire (OECD, 2013), whereas others were constructed by the authors.

[11]Providing extrinsic incentives to teachers based on student achievement has been found to have ambiguous results (e.g. Fryer (2013)).

3.3 Measuring Gender Role Beliefs

We measure the gender stereotypes of both students and teachers using the same questionnaire. This questionnaire includes a battery of item-set questions based on a 4-point Likert scale, with which we construct a "gender stereotype score" for each teacher and each student. Some examples are "It is more important for boys to go to college than girls", "Women cannot play football well even if they try hard", "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", which are to be answered using the scale of "I strongly agree", "I agree", "I disagree" and "I strongly disagree" (the full set of questions is given in the Appendix). Figure 1 shows the distribution of the stereotype scores of children and teachers, with larger numbers representing more traditional views. In both panels we see substantial variation in gender role beliefs, with male students and male teachers generally reporting more traditional views. For female teachers, we observe a clear pattern of piling up at the extremes (very progressive and very traditional) with considerable variation in between. For children, the distributions look fairly normal.

Table 1 presents the predictive power of teacher characteristics on the teacher's gender role beliefs. While male teachers seem to hold more stereotypical views about gender roles, this relationship does not reach statistical significance, and once teaching styles and effort are controlled for, it becomes even weaker. Years of experience have no bearing in predicting teachers' gender role beliefs. Several other interesting findings are noteworthy here. First, in terms of on-paper qualifications, teachers with a plain education (class teacher) degree are more likely to hold traditional beliefs about gender. This may be because this degree is less academically demanding and individuals who select into (or are placed because of their university entrance exam performance into) this major may be coming from a more traditional or less affluent background. Second, our teaching style constructs are by far the best predictors of teachers' gender role beliefs. Adding these constructs to the regression increases the R-square substantially (from 4% to 35%), and not surprisingly, a joint test of all style measures having no effect is decisively rejected. Among these style constructs,

growth mindset and warmth are the most important factors in determining teachers' gender role beliefs. Third, only one of our effort measures is statistically significant. Finally, the number of years taught in the same class does not predict teachers' gender role beliefs.[12] We now turn to estimating the effect of teachers' gender role beliefs on the achievement outcomes of students.

4 Results

While we were informed by school officials that the students are allocated to teachers within schools randomly, it is still useful to see whether our data attest to that. To do this, we look at the balance of fixed student and family characteristics across types of teachers. We construct two types to facilitate this balance check. Teachers with gender stereotype scores below the median are taken to be "progressive", while those with scores above the median are taken to be "traditional". While we use our continuous measure in our main analysis, this categorization also help us conduct a causal mediation analysis as detailed in Section 5. Table 2 presents the mean characteristics of students and families for traditional and progressive teachers. As can be seen clearly, all fixed student characteristics (including fluid IQ) and family characteristics that are unlikely to be affected by teachers' beliefs are balanced across the two types of teachers. The most notable evidence against the possibility of ability sorting is that our measure of IQ (elicited via Raven's progressive matrices) is balanced across the

[12]We also estimate a probability model for teaching the same class long-term. Table A.1 in the Online Appendix presents the results. Based on observable teacher characteristics, we do not find any consistent evidence suggesting that the teachers who taught the same class for a long term are a selected group. The only noteworthy exception is teachers with a linguistics degree. We find that they are about 0.53 percentage points more likely to stay in the same school for a long-time (p-value=0.01). We believe that this is due to the excess demand for teachers who can teach a foreign language in addition to regular class teaching in needy schools. These teachers are less likely to be rotated by the Ministry upon appointment.

two types of teachers.[13]

4.1 Empirical Specification

We use the following empirical model to estimate the effect of teachers' gender role beliefs on students' outcomes:

$$y_{iks} = cons + \alpha_1 Exposure_{iks} + \alpha_3 GRB_{ks} + \alpha_4 Exposure_{iks} * GRB_{ks} + \mathbf{X_{1,iks}}\beta + \mathbf{X_{2,iks}}\gamma + \mathbf{X_{3,ks}}\theta + \delta_s + \varepsilon_{iks}$$

where y_{iks} is the standardized test score for student i, who is being taught by teacher k in school s. The variable $Exposure$ captures the number of years student i has been taught by teacher k in school s. The variable GRB_{ks} is the continuous (standardized) score that measures the gender role beliefs of teacher k, with larger numbers representing more traditional beliefs. The interaction term allows for a differential effect of the teacher's beliefs on student outcomes with respect to the length of exposure to the teacher. Matrix $\mathbf{X_1}$ contains student characteristics such as age (in months), cognitive ability (as measured by the Raven IQ test), student's own gender role beliefs, student mindset, behavior score assigned by the teacher and an academic self-confidence measure. Matrix $\mathbf{X_2}$ contains family characteristics and socioeconomic indicators, and $\mathbf{X_3}$ contains teacher characteristics such as gender, experience as a teacher, education, study major, teaching styles and effort. Finally, δ_s denotes school fixed effects.

We divide the exposure variable into three groups: Children who have been taught by

[13]We also performed another check that involves predicting student achievement with only family socioeconomic indicators and looking at the correlation between the predicted values and teacher gender stereotype scores. If there is significant ability sorting, this correlation would be statistically significant. In both math and verbal and for both genders, we find no significant correlation between predicted test scores and teachers' gender views (p-values for math: girls=0.95, boys=0.16; for verbal: girls=0.39, boys=0.72). These findings provide supportive evidence that allocation of students to teachers is indeed random.

the participating teacher for at most one year are labeled as "1-year exposure", those who have been taught for more than one year and at most three years are labeled as "2-3 year exposure" and those who have been taught for more than three years (at most four years) are labeled as "4-year exposure".[14] As mentioned before, we have substantial variation in exposure due mainly to teacher relocation and to a lesser extent, family relocation.[15] Note that only fourth-grade students can be taught by the same teacher for more than three years in our sample, therefore our results regarding long-term exposure relates to fourth graders.

Given the random allocation of students to teachers, the (conditional) exogeneity of length of exposure, and the fact that we allow for school fixed effects, the coefficient estimates α, which are the estimates of interest, can be interpreted as causal effects. Despite our efforts of collecting very detailed information on teachers, we are cautious about the possibility that teacher gender role beliefs may still be capturing some unmeasured aspect of teacher quality. However, our i) gender-differential results presented in the next subsection, and ii) mediation analysis in Section 5 largely mitigate this concern.

4.2 Gender Role Beliefs of Teachers and Student Achievement

We estimate the empirical model presented in Section 4.1 separately for girls and boys. In addition to being of direct interest, looking at the effect of the beliefs separately for each gender also allows us to answer the question of whether beliefs still capture some unmeasured

[14]Because of the small sample size with respect to teachers in two-year exposure, we are not able to divide "2-3 year" further. We provide disaggregated estimation results in the Online Appendix (Figure A.1).

[15]About 13% of the students have been exposed to the same teacher less than their classmates. We consider them as re-locators. We were informed that the newcomers are allocated to classrooms in a random manner. Unreported regressions reveal that while they seem to be more likely to come from very low SES, their cognitive and non-cognitive skills, including their math and verbal test scores do not appear to be different from the rest of the sample. Results are available upon request.

aspect of teacher quality. If, although we control for many important teacher characteristics, beliefs still proxied teacher quality, we would expect to estimate similar effects on both genders unless such omitted characteristics have differential effects on boys and girls. We argue that this is unlikely to be the case and revisit the issue in Section 5. Table 3 presents the results by suppressing the coefficient estimates of student, family and teacher characteristics. Table A.2 in the Online Appendix gives the full results, which shows almost all cognitive and noncognitive ability measures we have are highly predictive of math and verbal test scores for both boys and girls. For math scores for example, a one standard deviation increase in the Raven (IQ) score is associated with 0.35 (0.23) standard deviations increase in math scores for girls (boys). Another important finding is that students' own gender role beliefs are also strong predictors of test scores for both genders: a one standard deviation increase in the gender stereotype score (going toward more traditional views) leads to about a 0.14 (0.12) standard deviation decrease in math scores girls (boys) and 0.12 standard deviation decrease in verbal scores for both boys and girls.

We now turn to the question of whether the teacher's beliefs affect girls' and boys' outcomes differently. What is clearly seen in Table 3 is that the teachers' gender role beliefs affect math and verbal test scores only for girls. The impact on math test scores is of considerable size, particularly when the girls have been taught by the same teacher for a long time (four years). A one standard deviation increase in teachers' gender stereotyped beliefs lowers girls' test scores in mathematics by about 0.21 standard deviations. The effect for an exposure of 2-3 years is smaller: a one standard deviation increase in teachers' gender stereotyped beliefs lowers girls' test scores in mathematics by about 0.12 standard deviations. While the equality of coefficients for 4-year and 2-3 year exposure is not rejected for either gender, we estimate a statistically significant effect of 4-year exposure to the same teacher relative to 1-year exposure for girls. No such effect is present for boys. Remarkably similar findings are obtained for the verbal scores (columns 3 and 4). Again, the impact of the teacher's stereotyped beliefs on girls' verbal test scores in the long term is of considerable

size (0.06 and 0.11 standard deviations for 2-3-year and 4-year exposure, respectively) and statistically significant at the 5% level.

When we test the effect of teachers' gender stereotypes for each exposure length across boys and girls, for math, we reject equality only for the 4-year exposure group (p-value=0.066) but for verbal, girls have a significant short-term advantage that is lost as they are exposed to the gender-biased teacher for a longer time. These results suggest that traditional gender role beliefs on the part of the teacher have a detrimental effect on girls' performance in both mathematics and verbal tests. However, the effects become visible after they spend some years with the same teacher. No such effect is present for boys. Finally, boys' math scores are significantly positively affected by long-term exposure to the same teacher, regardless of the teacher's gender role beliefs.[16] For girls in math, this relationship is weak and is reversed by being exposed to a teacher who holds traditional beliefs.

Since our measure of gender role beliefs is a continuous construct, it would be informative to present the functional relationship between test scores and teacher's beliefs in a non-parametric fashion. For this, we relax our assumption of a linear parametric model and modify our empirical model as follows:

$$y_{iks} = cons + \mathbf{X_{1,iks}}\beta + \mathbf{X_{2,iks}}\gamma + \mathbf{X_{3,ks}}\theta + \delta_s + f(GRB_{ks}) + \epsilon_{iks}$$

Here, while all student, family and teacher characteristics enter the model linearly, we allow for test scores to be a non-parametric function of the teacher's gender role beliefs (GRB). We estimate this model separately for boys and girls for each exposure length. Recall that larger numbers of GRB indicate more traditional (stereotyped) beliefs. Figure 2 depicts the results for math test scores. Our findings from the linear models clearly re-emerge

[16]Related to this result, Hill and Jones (2017) find that repeat student-teacher matches have a significantly positive effect on student achievement in similar (3rd to 5th grade) elementary school students, pointing to the benefit of staying with the same teacher.

for girls in these pictures. Looking at 4-year and 2-3 year exposure results, one can see the decreasing and fairly linear relationship between the gender stereotypes of teachers and girls' math test scores. For boys on the other hand, we observe a rather non-linear relationship where at the very extreme (most progressive teachers) they exhibit similar patterns as girls: boys' math scores are higher under extremely progressive teachers, however the relationship breaks down as the teacher becomes more conservative. It appears that except for the case of an extremely progressive teacher, boys may even be benefiting from a teacher's traditional gender role beliefs (note the slight positive relationship, not considering the extremes). For 1-year exposure, the relationship is virtually flat for both boys and girls, with again some evidence of both genders benefiting from a very progressive teacher.

As for the verbal scores, Figure 3 depicts the negative functional relationship between the teacher's beliefs and girls' verbal test scores. With again the exception at the corner (most progressive teachers), the relationship is flat for boys. Overall, our results suggest a significantly gender-differential effect of the teacher's gender role beliefs on student achievement. Under both parametric and non-parametric specifications, we estimate a declining and fairly linear relationship for girls under 4-year exposure to the same teacher, while no obvious (statistically significant) pattern of relationship emerges for boys. We now turn to investigate the sensitivity of our results to various issues raised earlier.

4.3 Robustness

The behavior at the extreme (very progressive teachers) is noteworthy. Given the similar (positive) effects of such teachers on the test scores of both boys and girls, it may be that some omitted aspects of teacher quality are proxied well with extreme progressiveness. In Table A.3, we re-estimate Table 3 by excluding very progressive teachers, in order to see how sensitive our results are to these particular teachers. For this, we exclude teachers whose gender stereotype score is lower than the 10th percentile (15 teachers, two of them male). As can be seen in the table, the results for girls, especially for math scores remain very strong,

although we lose some precision for verbal results.

Even though our identification relies on within-school variation through the use of school fixed effects, we conduct another robustness check that is related to teacher sorting into schools. Recall that the institutional structure leaves very little room for self-selection of teachers into catchment areas/schools and our sample consists of generally "undesirable" schools. However, although still difficult, teachers who have accumulated high service points (those with higher number of years of service) might be able to self-select into relatively more desirable schools. Given that working in a catchment area of one's choice is generally ruled out before 20 years of service except purely by chance, we re-estimate our linear model by excluding the teachers who have more than 20 years of service in the teaching profession. This excludes 24 teachers from our sample. Table A.4 presents the results for boys and girls separately. Results are both qualitatively and quantitatively very similar to our full sample results.

Another concern one might have is that a teacher's beliefs may reflect what he/she observes in the class. Suppose that in a given classroom or cohort, boys are indeed better academically than girls. If the teacher bases his/her beliefs on this particular cohort, our results would reflect this reverse causality rather than the effect of the teacher's beliefs on achievement. Our rich data, however, allows us to address this issue. Our teacher survey includes a question where we ask the teacher whether he/she has observed boys or girls to be better at math (or equal) in his/her experience as a teacher. When we exclude the teachers who report boys to be better (only 7 teachers), our results remain the same (see Table A.5).[17]

Responses to the question of which gender tends to be better at math also reveal that the teachers in our sample do not maintain stereotyped beliefs about mathematical ability across gender. 56% of our teachers report that they have observed girls to be better at math

[17]Our results also hold when we entirely exclude this question from our gender role belief construct and base the measure on other domains of gender stereotypes than math performance.

and about 39% report that both genders are equally good, with only about 5% thinking boys are better.[18] The lack of a stereotype about math ability is also evident in our findings regarding grades. As can be seen in Table 4, we observe absolutely no effect of teachers' gender role beliefs on students' grades. The absence of an effect on grades suggests that the effects we estimate on objective achievement scores do not reflect reverse causality, i.e., they are not coming from teachers' factual beliefs about ability (based on their observations over the years or in their current classroom).[19] In the next section, we explore a potential mechanism that may lead to these results.

5 A Causal Mediation Analysis

Recall that Table A.2 shows that various student characteristics, which may be affected by teachers' gender role beliefs, are highly predictive of test scores and therefore may be potential mediators of the effects we estimate. An obvious one is students' own gender role beliefs. If girls adopt the biased beliefs held by their teacher, this may diminish their ambitions, aspirations, and motivation towards academic tasks, reducing their achievement. Another mediator may be self-confidence. Our measure of self-confidence is derived from a survey item designed to measure students' beliefs on their math performance ("In math, I

[18]In our data, the unconditional performance of girls and boys in objective math test is similar; however, the dummy for male becomes strongly and positively significant in explaining math performance once we control for other student characteristics. As for verbal performance, the unconditional performance of girls is significantly higher but this advantage turns statistically insignificant once we control for student characteristics. All these hold true for math and verbal grades as well.

[19]The absence of an effect on grades despite the effect on objective tests may also point to the fact that grades tend to reflect non-cognitive skills and good behavior in addition to pure exam performance, especially in elementary school (e.g. Brookhart (1993), McMillan et al. (2002), Borghans et al. (2016), Jackson (2016)). Such effects may also potentially explain findings of grading biases in favor of girls (e.g. Terrier (2015)).

am: very good/good/mediocre/not very good/not good at all"). A traditional teacher may potentially affect girls' confidence in mathematics by either directly voicing beliefs about girls' capabilities or praising/focusing on boys more in math. Finally, another potential mediator could be the students' mindset on achievement, i.e. whether students have a "growth mindset" that highlights the importance of effort or a "fixed mindset" that emphasizes innate abilities. Gender-biased teachers, who hold fixed views of what each gender can and cannot do, may influence the achievement mindset of students, particularly girls. This shift towards a fixed mindset may in turn lead to lower motivation and performance, as has been shown in the literature (Blackwell et al. (2007), Alan, Boneva and Ertac (2016)).

In addition to these indirect channels, or alternatively, teachers' gender role beliefs may affect student achievement directly. A teacher with strongly traditional gender role beliefs, who thinks that it is more important to get boys to do well in school, may adopt classroom practices that reflect these beliefs; for example, asking questions to and answering questions from girls and boys differently, providing more feedback to boys, and generally focusing academic attention more on boys while praising girls for gender-consistent behavior such as compliance and obedience (Dweck et al. (1978)). These practices may impede girls' learning directly, without necessarily affecting their own gender role beliefs.

In order to establish whether and how large a part of the effect on test scores is coming through these potential mediators, we perform a statistical mediation analysis. For this, we use an extension of the potential outcomes framework developed by Imai et al. (2010) to estimate causal mediation effects. To make the analysis feasible and facilitate straightforward interpretation, we use a binary teacher gender stereotype score to serve as a binary treatment indicator. Teachers with scores below the median are taken as "progressive", while those with scores above the median are taken as "traditional".[20] Recall that conditional on school, being exposed to a particular type of teacher is random in our setting.

[20]Doing this analysis with a continuous treatment variable is not trivial. Also the interpretation of the results would be very difficult.

While the random assignment to a type of teacher is sufficient to identify the total effect, additional (strong) assumptions are required to identify the average causal mediation effect (ACME) and the average direct effect (ADE). Imai et al. (2010) show that ACME and ADE can be nonparametrically identified under the "sequential ignorability" assumption, which constitutes two sequential conditions. The first one states that given the pre-treatment confounders, treatment assignment is independent of the potential outcomes and potential mediators. The second one states that the mediators are independent of the potential outcomes conditional on pre-treatment confounders and the treatment assignment. While we make use of our rich data on numerous student, family and teacher characteristics that potentially affect both the mediators and the outcome, the latter is still a very strong assumption.

To estimate the average effects (ACME and ADE), we proceed in several steps. First, we posit and fit regression models for the mediator (say, students' own gender role beliefs) and the outcome of interest (test scores). The mediator model includes the treatment dummy (traditional teacher) as well as any relevant covariates. The outcome is modeled as a function of the mediator and the treatment dummy, as well as all covariates. Based on the fitted mediator model, we then generate two sets of predicted mediator values for each girl, one under a progressive teacher and the other under a traditional teacher.

We then use the outcome model to impute potential outcomes. For each girl, we first obtain the predicted value of the outcome corresponding to the traditional teacher and the predicted mediator value for the treatment condition (obtained in the previous step). We then generate the predicted counterfactual outcome, i.e. the outcome where the treatment indicator is still set to 1 (traditional teacher) but the mediator is set to its predicted value under the progressive teacher (also obtained in the previous step). Finally, we compute the average causal mediator effect by averaging the differences between the predicted outcome under the two values of the mediator across observations in the data.

Table 5 presents the effects of teacher beliefs on the three potential mediators we consider. Pooling all exposure lengths, we estimate that a traditional teacher increases girls' gender

stereotyped beliefs by about 0.20 standard deviations (p-value=0.001). The relationship is not statistically different from zero for boys (p-value=0.96). We estimate no effect on self-confidence and mindset for either boys or girls. Table 6 presents the average causal mediation effect (ACME), average direct effect (ADE) and total effect for both math and verbal test scores for each gender. Overall, we estimate that a traditional teacher lowers girls' math scores by about 0.16 standard deviations. About a 0.03 standard deviation of that (17%) comes from girls' gender role beliefs being affected by their teacher's gender role beliefs. The rest of the effect, not mediated by beliefs, may be due to the direct effect of factors such as lower academic attention on girls by traditional teachers.[21]

The results on verbal scores is quite interesting. The total effect of the teacher's gender role beliefs on verbal performance is not statistically different from zero in this specification; however, transmission of the teacher's gender role beliefs to female students leads to an approximately 0.02 standard deviation decline in verbal scores, making the total effect smaller than ADE. Both ACME estimates (math and verbal) are statistically significant (see the 95% confidence intervals). It should be noted here that these numbers are just direct effects of level shifts in gender role beliefs. It is quite possible that changes in these beliefs affect performance through indirect influences on girls' perceived or true production function. For example, a girl who holds biased beliefs may have lower motivation in a mathematical performance task. As expected, all estimates are not statistically different from zero for boys.[22]

Note that our analysis show that self-confidence is not a potential channel. This finding along with the finding of a significant effect that is mediated by girls' gender role beliefs points to the role of potential indoctrination about what is expected of a woman, which

[21]When we exclude short-term exposure (as we find no effect in this case), we lose considerable precision in the mediator model and this results in a lower percentage (about 14%) of the total effect being mediated.

[22]We also performed this analysis using gender roles in the family as a potential mediator and ruled it out. Results are available upon request.

may lower girls' academic motivation or ambitions. That is, rather than lowering girls' self-confidence about their capabilities, traditional teachers may emphasize appropriate roles for them in the society. If traditional teachers emphasize traditional gender roles whereby girls do not need to be as ambitious as boys in the academic domain (because they won't need to use these skills as much), this may manifest in lower academic motivation in girls, although their beliefs about their capabilities do not necessarily go down. In fact, the set of questions in our survey about appropriate gender roles (e.g. the proper division of labor within the family) are responsible for the effect that comes from student beliefs. Among those, item set questions such as "it is the father's responsibility to earn money for the household", "it is natural for girls to help more than boys in household chores" are highly strong mediators when considered in isolation. We should note that the traditional teacher may also place less academic attention on girls, which may have a strong "direct" effect on their learning that is not mediated through student beliefs.

One alternative explanation of our differential results across gender would be a differential response of girls and boys to teaching quality.[23] Although we have a large set of controls for teacher characteristics, if gender role beliefs still capture an unmeasured aspect of teacher quality and girls' achievement is more responsive to this, similar patterns would emerge. Our data, however, provides suggestive evidence against this. Table A.2 shows that boys' achievement is at least as responsive to teacher characteristics as girls'. Coefficient estimates on teacher characteristics do not suggest that girls are in any way more responsive to quality, styles and approach. Along with the result that teacher gender role beliefs are transmitted to girls more strongly, these results give us confidence that our findings are coming from the teacher acting on biased gender views and conveying these beliefs to children, rather than an unmeasured aspect of teacher quality (correlated with teacher gender role beliefs) affecting girls differentially.

[23]Deming et al. (2014) show that at the high-school level, girls respond to attending a better school with higher grades and taking more courses to prepare for college.

6 Conclusion

We exploit a natural experiment to show that teachers' gender role beliefs have a significant impact on girls' math and verbal test scores. Our unique setting allows us to identify the effects moderated by the duration of teacher contact with students. Controlling for student, family and teacher characteristics, we show that girls whose teachers maintain more traditional (progressive) views about gender roles have lower (higher) performance in objective math and verbal tests, and this effect is amplified with longer exposure to the same teacher. For boys, we find no significant effect.

The large dataset we use, collected with the purpose of answering the research question we pose in this paper, allows us to control for a host of teacher, student and family characteristics that are crucial for identifying the effect of gender role beliefs on achievement. The results show that controlling for the teacher's own gender and other characteristics, teachers' beliefs about gender roles affect the test scores of their female students, both in mathematics and verbal tests. It is striking that even without any apparent biases or discrimination in grading, teachers' traditional gender role beliefs still affect girls' achievement outcomes negatively. Our mediation analyses show that a non-trivial portion of the effect comes from the teacher transmitting his/her traditional gender role beliefs to girls. These results indicate that the personal views of the elementary school teacher may play an important role in mitigating or widening gender-achievement gaps, particularly in countries where pervasive gender inequality has been found to contribute to differences in math performance across gender (Guiso et al. (2008)). Given that our sample comes from the low socioeconomic tier, our results are also generalizable to vulnerable segments of societies, where patriarchal gender roles are particularly imposing and where improving achievement is a policy imperative (Heckman (2006)).

Two caveats are worth mentioning. First is the fact that our data is a cross section. Panel data with some baseline information on students before they were exposed to a particular teacher would of course be ideal, especially to pin down heterogeneous effects of teacher

types. The second one is the external validity of our results. To circumvent the issue of ability sorting of students, we exploit our unique country setting and choose our sample from lower socioeconomic strata (relatively deprived areas of Istanbul). In this group, teachers are prevented from self-selecting into schools and students are randomly allocated to teachers. While giving us a clean identification of the effects of teacher types on achievement, this choice may prevent us from generalizing our findings to the population. Future work should focus especially on these two issues.

Given the importance of the childhood period for long-term choices and outcomes, the results suggest that the type of teacher a child is assigned to in elementary school may have long-lasting consequences. In particular, improved math scores of girls may lead to reductions in gender gaps in the labor market, given the evidence that math performance and math education predict future income (Paglin and Rufolo (1990), Joensen and Nielsen (2009)). The implication for educational policy is that achieving gender equality in teaching practices and attitudes early on, possibly by training teachers to raise awareness of such biases and their effects, could have substantial value for preventing inefficient gender gaps in achievement, occupational selection and labor market outcomes.

References

[1] Aaronson, Daniel and Barrow, Lisa and Sander, William (2007). "Teachers and student achievement in the Chicago public high schools." *Journal of Labor Economics*, 25(1), 95-135.

[2] Alan, Sule and Boneva, Teodora and Ertac, Seda (2016), "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit", HCEO Working Paper.

[3] Almlund, Mathilde, Duckworth, Angela Lee and Heckman, James and Kautz, Tim. (2011). "Personality Psychology and Economics." *Handbook of the Economics of Education*, 1–181.

[4] Antecol, Heather and Eren, Ozkan and Ozbeklik, Serkan. (2014). "The effect of teacher gender on student achievement in primary school." *Journal of Labor Economics*, 33(1), 63-89.

[5] Bertrand, Marianne. (2011). "New perspectives on gender." *Handbook of Labor Economics*, 4, pp. 1543-1590.

[6] Bertrand, Marianne and Hallock, Kevin F. (2001). "The gender gap in top corporate jobs". *Industrial & Labor Relations Review*, 55(1), 3-21.

[7] Bettinger, Eric P and Long, Bridget Terry. (2005). "Do faculty serve as role models? The impact of instructor gender on female students." *The American Economic Review*, 95(2), 152-157.

[8] Bietenbeck, Jan. (2014). "Teaching practices and cognitive skills." *Labour Economics*, 30, 143-153.

[9] Bisin, Alberto and Verdier, Thierry. (2001). "The Economics of Cultural Transmission and the Dynamics of Preferences." *Journal of Economic Theory*, 97: 298-319.

[10] Blackwell, Lisa S and Trzesniewski, Kali H and Dweck, Carol Sorich. (2007). "Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention." *Child Development*, 78(1), 246-263.

[11] Blau Francine, D and Ferber, Marianne A and Winkler, Anne E. (2002), *The Economics of Women, Men and Work* (Englewood Cliffs, NJ: Prentice Hall).

[12] Borghans, Lex and Golsteyn, Bart HH and Heckman, James J and Humphries, John Eric. (2016). "What grades and achievement tests measure." *Proceedings of the National Academy of Sciences*, 113(47), 13354-13359.

[13] Brookhart, Susan M. 1993. "Teachers' Grading Practices: Meaning and Values." Journal of Educational Measurement 30(2): 123 –42.

[14] Carrell, Scott E and Page, Marianne E and West, James E. (2010). "Sex and science: How professor gender perpetuates the gender gap." *The Quarterly Journal of Economics*, 125(3), 1101-1144.

[15] Carrell, Scott E and West, James E. (2010). "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy*, 118(3), 409-432.

[16] Chetty, Raj and Friedman, John N and Rockoff, Jonah E. (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *The American Economic Review*, 104(9), 2633-2679.

[17] Dee, Thomas S. (2005). "A teacher like me: Does race, ethnicity, or gender matter?." *The American Economic Review*, 95(2), 158-165.

[18] Dee, Thomas S. (2007). "Teachers and the gender gaps in student achievement." *Journal of Human Resources*, 42(3), 528-554.

[19] Deming, David J and Hastings, Justine S and Kane, Thomas J and Staiger, Douglas O. (2014). "School choice, school quality, and postsecondary attainment." *The American Economic Review*, 104(3), 991-1013.

[20] Domino, George. (1971). "Interactive effects of achievement orientation and teaching style on academic achievement." *Journal of Educational Psychology*, 62(5), 427.

[21] Dweck, Carol S. (2006). Mindset: The new psychology of success. Random House Incorporated.

[22] Dweck, Carol S and Davidson, William and Nelson, Sharon and Enna, Bradley. (1978). "Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis." *Developmental Psychology*, 14(3), 268.

[23] Farre, Lidia and Vella, Francis. (2013). "The intergenerational transmission of gender role attitudes and its implications for female labour force participation." *Economica*, 80(318), 219-247.

[24] Fortin, Nicole M. (2005). "Gender role attitudes and the labour-market outcomes of women across OECD countries." *Oxford Review of Economic Policy*, 21(3), 416-438.

[25] Fryer, Roland G. (2013). "Teacher incentives and student achievement: Evidence from New York City public schools." *Journal of Labor Economics*, 31(2), 373-407.

[26] Guiso, Luigi and Monte, Ferdinando and Sapienza, Paola and Zingales, Luigi. (2008). "Culture, Gender, and Math." *Science*, 320, No. 5880 , 1164 - 1165.

[27] Hanushek, Eric and Rivkin, Steven. (2006). "Teacher quality." *Handbook of the Economics of Education*, 2, 1051-1078.

[28] Hanushek, Eric A. (2011). "The Economic Value of Higher Teacher Quality." *Economics of Education Review*, 30, 466-479.

[29] Harris, Douglas N and Sass, Tim R. (2011). "Teacher training, teacher quality and student achievement." Journal of public economics, 95(7), 798-812.

[30] Heckman, James J. (2006). "Skill formation and the economics of investing in disadvantaged children." Science, 312, No. 5782, 1900-1902.

[31] Hidalgo Cabrillana, Ana and Lopez-Mayan, Cristina. (2015). "Teaching styles and achievement: Student and teacher perspectives." Working paper.

[32] Hill, Andrew J. and Jones, Daniel B. (2017). "A teacher who knows me: The academic benefits of repeat student-teacher matches." Working paper.

[33] Hoffmann, Florian and Oreopoulos, Philip. (2009). "A professor like me: the influence of instructor gender on college achievement." *Journal of Human Resources*, 44(2), 479-494.

[34] Imai, Kosuke and Keele, Luke and Yamamoto, Teppei. (2010). "Identification, inference and sensitivity analysis for causal mediation effects." *Statistical Science*, 25(1), 51-71.

[35] Jackson, C Kirabo. (2016). What do test scores miss? The importance of teacher effects on non-test score outcomes (No. w22226). National Bureau of Economic Research.

[36] Joensen, Juanna Schrøter and Nielsen, Helena Skyt. (2009). "Is There a Causal Effect of High School Math on Labor Market Outcomes?" *Journal of Human Resources* 44(1): 171-198.

[37] Johnston, David W and Schurer, Stefanie and Shields, Michael A. (2014). "Maternal gender role attitudes, human capital investment, and labour supply of sons and daughters." *Oxford Economic Papers*, 66(3), 631-659.

[38] Lavy, Victor. (2008). "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of Public Economics*, 92(10), 2083-2105.

[39] Lavy, Victor and Sand, Edith. (2015). "On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases" (No. w20909). National Bureau of Economic Research.

[40] McMillan, James H and Myran, Steve and Workman, Daryl. (2002). "Elementary teachers' classroom assessment and grading practices." *The Journal of Educational Research*, 95(4), 203-213.

[41] Olivetti, Claudia and Patacchini, Eleonora and Zenou, Yves. (2016). "Mothers, Peers and Gender Identity." Working Paper.

[42] OECD (2013). TALIS User Guide for the International Database, http://www.oecd.org/edu/school/TALIS-2013-User-guide.pdf

[43] Paglin, Morton and Rufolo, Anthony M. (1990). "Heterogeneous human capital, occupational choice, and male-female earnings differences." *Journal of Labor Economics*, 8(1, Part 1), 123-144.

[44] Rivkin, Steven G and Hanushek, Eric A and Kain, John F. (2005). "Teachers, schools, and academic achievement." *Econometrica*, 73(2):417–458.

[45] Robinson, Joseph Paul and Lubienski, Sarah Theule. (2011). "The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings." *American Educational Research Journal*, 48(2), 268-302.

[46] Rockoff, Jonah E. (2004). "The impact of individual teachers on student achievement: Evidence from panel data." *The American Economic Review*, 94(2), 247-252.

[47] Sadker, Myra and Sadker, David. (2010). "Failing at fairness: How America's schools cheat girls." Simon and Schuster.

[48] Schwerdt, Guido and Wuppermann, Amelie C. (2011). "Is traditional teaching really all that bad? A within-student between-subject approach." *Economics of Education Review*, 30(2), 365-379.

[49] Spencer, Steven J and Steele, Claude M and Quinn, Diane M. (1999). "Stereotype threat and women's math performance." *Journal of Experimental Social Psychology*, 35(1), 4-28.

[50] Terrier, Camille. (2015). "Giving a little help to girls? Evidence on grade discrimination and its effect on students' achievement." CEP Discussion Papers, CEPDP1341. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Table 1: Predictors of Teachers' Gender Role Beliefs

Male	0.241	0.251	0.269	0.101	0.124
	(0.18)	(0.18)	(0.19)	(0.17)	(0.17)
University Degree		-0.167	-0.173	-0.168	-0.171
		(0.31)	(0.32)	(0.31)	(0.29)
Graduate Degree		-0.506	-0.503	-0.457	-0.470
		(0.42)	(0.42)	(0.38)	(0.37)
Years of Experience		-0.007	-0.007	0.004	0.004
		(0.01)	(0.01)	(0.01)	(0.01)
Number of Terms in the Same Class		0.028	0.029	0.015	0.029
		(0.04)	(0.04)	(0.03)	(0.03)
Education Degree			0.336	0.409	0.595***
			(0.34)	(0.25)	(0.21)
Linguistics			-0.203	-0.271	-0.188
			(0.39)	(0.33)	(0.32)
Natural Sciences			0.271	0.141	0.230
			(0.40)	(0.19)	(0.23)
Social Sciences			-0.149	-0.235	-0.191
			(0.30)	(0.23)	(0.24)
Growth Mindset				-0.188***	-0.181***
				(0.03)	(0.03)
Extrinsic Motivator				0.033	0.016
				(0.04)	(0.04)
Modern Approach				-0.013	0.000
				(0.03)	(0.02)
Warm Approach				-0.101***	-0.096***
				(0.03)	(0.03)
Number of Extra_C Programs					0.005
					(0.01)
Number of Volunteer Activities					-0.041***
					(0.01)
N	145	145	145	145	145
R-Squared	0.01	0.03	0.04	0.35	0.39

The dependent variable is the teacher's standardized gender stereotype score. It is constructed in a way that larger values indicate more traditional gender role beliefs. Heteroskedasticity-robust standard errors are in parentheses.

Table 2: Balance Across Teacher Types

	Fixed Student Characteristics		
	Progressive	Traditional	P-Value
Male student	0.51	0.51	0.862
Age (in months)	109.5	109.8	0.739
IQ (Raven Score)	0.09	0.07	0.628
	Family Socioeconomic Indicators		
	Progressive	Traditional	P-Value
Working mother	0.30	0.26	0.130
Computer at home	0.75	0.75	0.675
Family gender roles	2.30	2.31	0.813
Low SES	0.34	0.36	0.760
Medium SES	0.44	0.43	0.847
High SES	0.22	0.21	0.845

The table presents mean values of fixed student characteristics (upper panel) and family socioeconomic indicators for progressive and traditional teachers. Progressive (traditional) teachers are defined as those whose gender role beliefs are below (above) the median score. IQ is measured (and standardized to have mean zero and variance 1) via Raven's Progressive Matrices. Binary indicators of whether the mother is working, whether there is a computer at home, and gender roles in the family are reported by the child. The latter is a question based on a 4-item scale that asks how much the father takes part in household chores. Family income/wealth level (SES) is reported by the teacher based on a 1-5 item scale and low, medium and high SES indicators are constructed based on these.

Table 3: Heterogeneous Effects of Teacher Gender Role Beliefs on Test Scores

	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)
2-3 Year Exposure	0.022	0.058	0.026	0.033
	(0.07)	(0.06)	(0.07)	(0.08)
4 Year Exposure	0.117	0.193***	0.015	0.006
	(0.08)	(0.07)	(0.08)	(0.07)
2-3 Year Exposure*Teacher G-Styping	-0.120*	0.001	-0.110**	0.020
	(0.06)	(0.06)	(0.05)	(0.07)
4 Year Exposure*Teacher G-Styping	-0.211**	-0.016	-0.162**	-0.026
	(0.08)	(0.07)	(0.07)	(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	0.229	0.792	0.428	0.442
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.067	0.992	0.037	0.764
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.014	0.831	0.032	0.719
P-value: 1 Year E*G-Sty[Girls=Boys]	0.393		0.004	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.157		0.689	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.066		0.834	
N	1870	1943	1873	1946
R-Squared	0.32	0.34	0.26	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.

Table 4: Heterogeneous Effects of Teacher Gender Role Beliefs on Grades

	Math Grade		Verbal Grade	
	Girls	Boys	Girls	Boys
Teacher G-Styping	0.060	0.082	0.144**	0.158*
	(0.10)	(0.12)	(0.07)	(0.09)
2-3 Year Exposure	0.035	0.134	0.094	0.141
	(0.10)	(0.11)	(0.09)	(0.10)
4 Year Exposure	-0.058	0.061	0.003	-0.020
	(0.11)	(0.12)	(0.10)	(0.11)
2-3 Year Exposure*Teacher G-Styping	0.069	0.004	-0.057	-0.100
	(0.10)	(0.12)	(0.06)	(0.10)
4 Year Exposure*Teacher G-Styping	0.028	-0.026	-0.039	-0.053
	(0.12)	(0.14)	(0.09)	(0.11)
School Fixed Effects	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓
P-value: 2-3 Year E*G-Styp=Long*G-Styp	0.536	0.681	0.770	0.525
P-value: 1 Year E*G-Styp=2-3 Year E*G-Styp	0.506	0.977	0.368	0.314
P-value: 1 Year E*G-Styp=4 Year E*G-Styp	0.813	0.852	0.670	0.644
P-value: 1 Year E*G-Sty[Girls=Boys]	0.808		0.852	
P-value: 2-3 Year E*G-Styp[Girls=Boys]	0.454		0.535	
P-value: 4 Year E*G-Sty[Girls=Boys]	0.672		0.988	
N	1594	1652	1594	1652
R-Squared	0.42	0.45	0.37	0.42

Dependent variables are standardized grades given by the teacher. 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 5: Mediator Model: The Effect of Teachers' Beliefs on Students' Beliefs

	Gender Role Beliefs		Self Confidence		Growth Mindset	
	Girls	Boys	Girls	Boys	Girls	Boys
Traditional Teacher	0.199***	-0.003	0.091	0.080	-0.003	-0.074
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.07)
School Fixed Effects	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓
Family Characteristics	✓	✓	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓	✓	✓
Teaching Styles	✓	✓	✓	✓	✓	✓
Teacher Effort	✓	✓	✓	✓	✓	✓
N	1888	1967	1888	1967	1888	1967

Dependent variables are standardized scores of students' gender role beliefs, self-confidence and growth mindset. The binary variable "Traditional Teacher" takes the value 1 if the teacher's beliefs are above the median score and zero otherwise. 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 6: Potential Channels for the Effects on Test Scores: Causal Mediation

PANEL 1: Math Test Scores			
	G. Role Beliefs	Self-confidence	Growth Mindset
ACME	**-0.028**	0.011	-0.000
	[-0.046, -0.011]	[-0.004, 0.028]	[-0.004, 0.003]
ADE	-0.136	-0.136	-0.136
	[-0.241, -0.021]	[-0.241, -0.021]	[-0.241, -0.021]
TOTAL	-0.164	-0.125	-0.136
	[-0.269, -0.045]	[-0.232, 0.008]	[-0.241, -0.021]
Percentage Mediated (%)	**17%****	-8.5%	0.12%

PANEL 2: Verbal Test Scores			
	G. Role Beliefs	Self-confidence	Growth Mindset
ACME	**-0.024**	0.006	-0.001
	[-0.041, -0.010]	[-0.002, 0.019]	[-0.011, 0.010]
ADE	0.048	0.048	0.048
	[-0.060, 0.165]	[-0.060, 0.165]	[-0.060, 0.165]
TOTAL	0.024	0.054	0.048
	[-0.083, 0.144]	[-0.053, 0.172]	[-0.059, 0.165]
Percentage Mediated %	30%	9.3%	-0.32%

ACME: Average causal mediation effect, ADE: Average direct effect. G.R.: Gender role. Estimates (standard deviation effects) and 95% confidence intervals are obtained via Imai et al. (2010). The estimation sample is restricted to girls only. Number of simulations is 1000. **: significant at 5%.

Figure 1: Distribution of Gender Role Beliefs

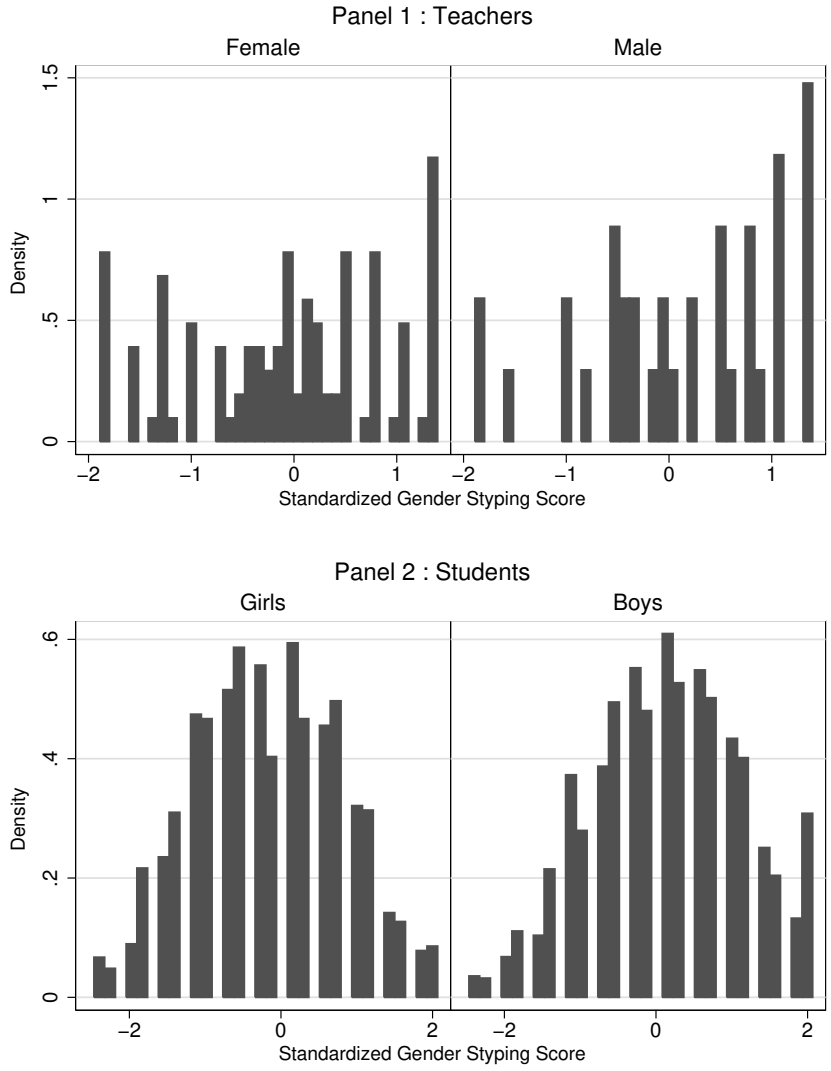
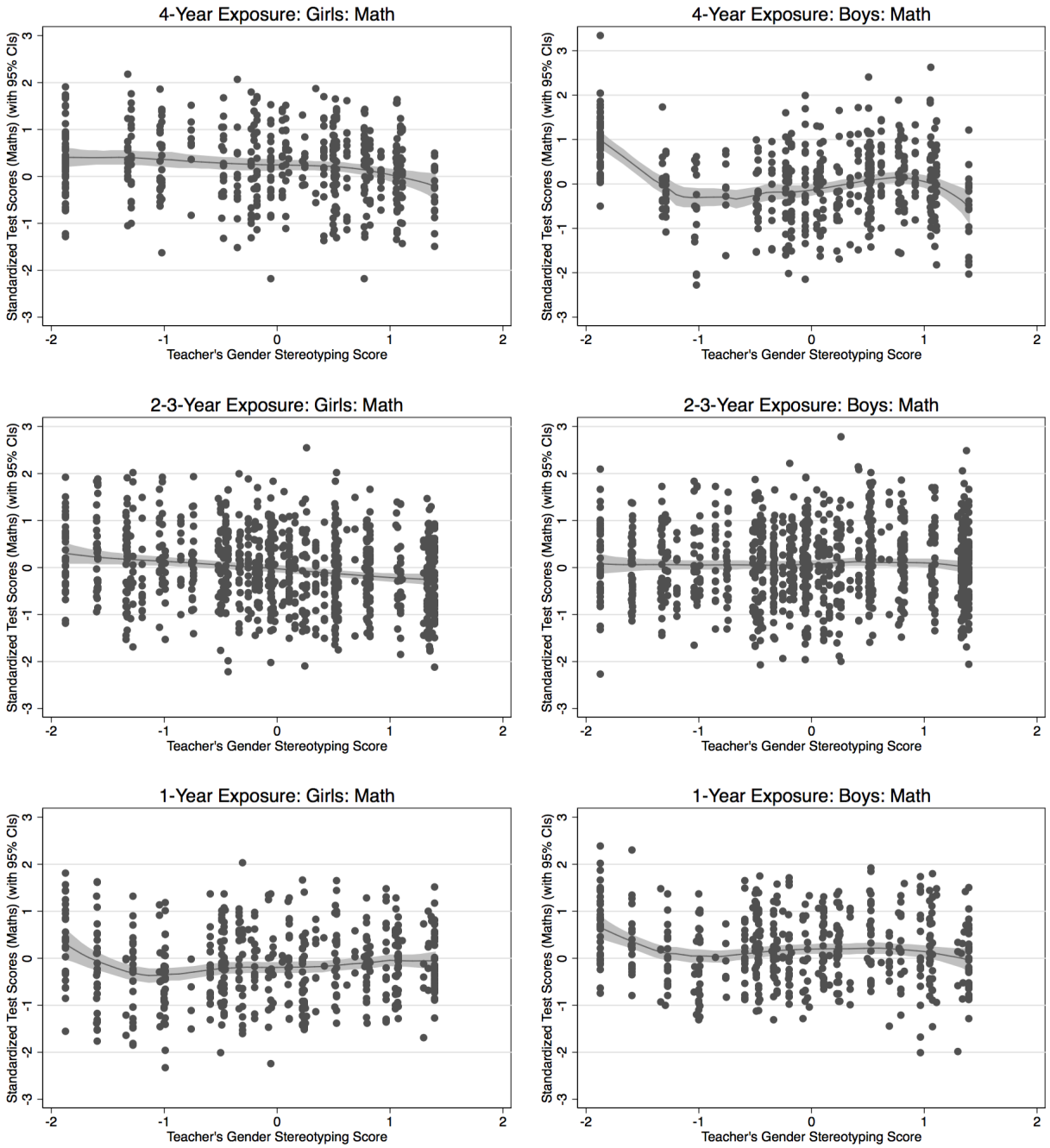
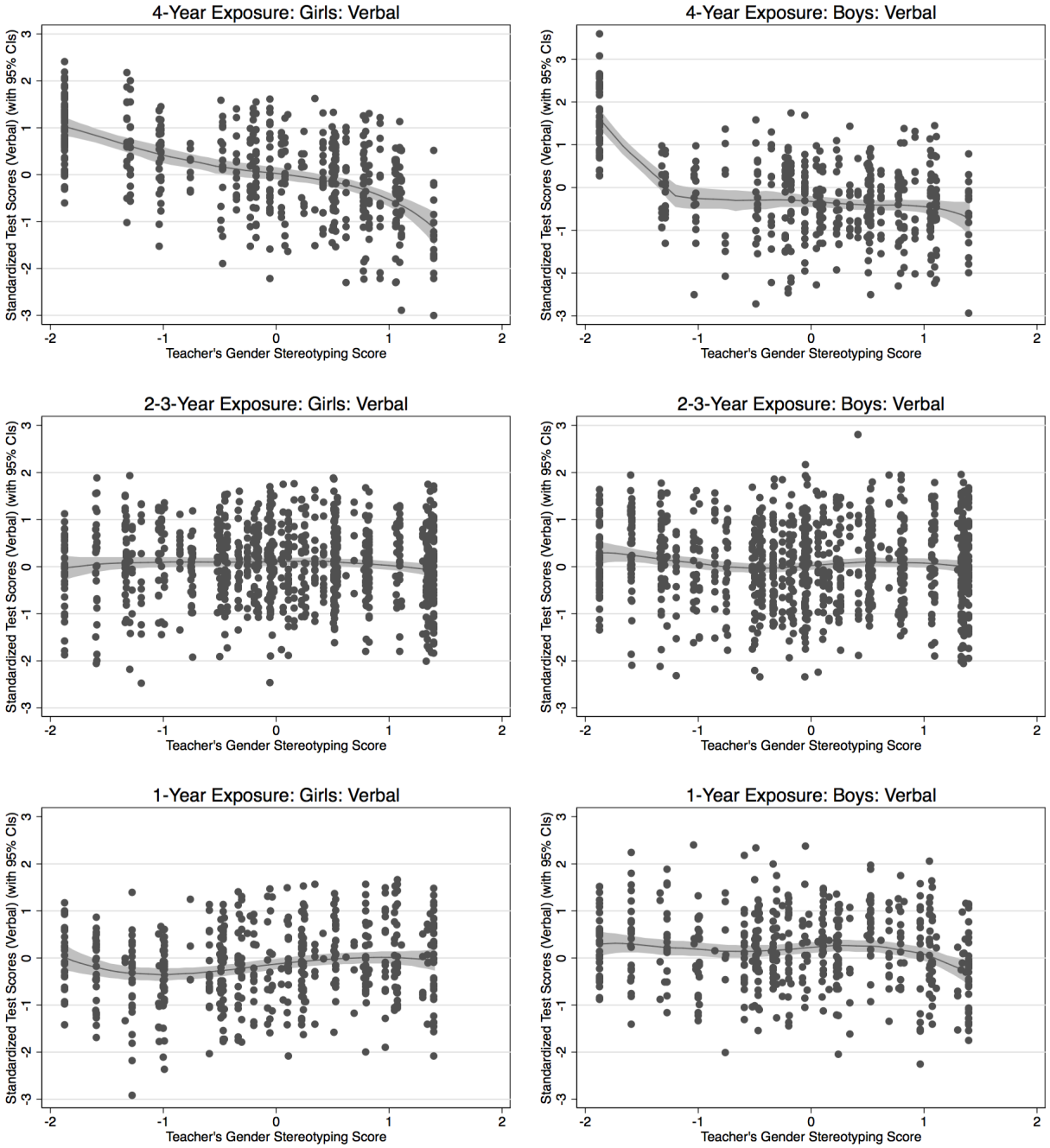


Figure 2: Teacher Gender Stereotyping and Math Test Scores: Non-parametric



Figures plot the non-parametric estimates (and 95% confidence bands) of the effect of teacher's role beliefs on math test scores for girls (column 1) and for boys (column 2). All student, family and teacher characteristics enter the model linearly, and school fixed effects are included.

Figure 3: Teacher Gender Stereotyping and Verbal Test Scores: Non-parametric



Figures plot the non-parametric estimates (and 95% confidence bands) of the effect of teacher's role beliefs on verbal test scores for girls (column 1) and for boys (column 2). All student, family and teacher characteristics enter the model linearly, and school fixed effects are included.

View publication stats