GENDER STEREOTYPES IN THE CLASSROOM AND EFFECTS ON ACHIEVEMENT

Sule Alan, Seda Ertac, and Ipek Mumcu*

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.

I. 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 leadership. As ample evidence suggests, such gender inequality, factually confirming and perpetuating traditional gender role beliefs, can be quite persistent (Bertrand & Hallock, 2001; Blau, Ferber, & 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 & 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 important, with another adult: the teacher. Teachers' views toward gender roles may affect students' attitudes, behaviors, and outcomes directly and indirectly. First, a teacher's beliefs may influence students' achievement outcomes by influencing students' own beliefs: the teacher may express his or her views about gender-appropriate roles in the classroom, and because he or she is a significant authority figure, students may adopt and internalize what that 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, Steele, & Quinn, 1999).

A more direct mechanism is the teacher's 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 or dismissing questions, or focusing on boys when teaching (Sadker & 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 & 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' noncognitive skills as well.[2] A very progressive teacher, in contrast, 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 4,000 third- and fourth-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 a natural experiment with three main components: First, stratified by gender and preschool education, state elementary school students are allocated to their teacher in first grade randomly. Second, teachers are appointed to schools centrally by the Ministry of Education based on the need for teachers and thus 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 1 to 4, 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

Received for publication August 2, 2017. Revision accepted for publication January 19, 2018. Editor: Amitabh Chandra.

* Alan: University of Essex; Ertac: Koc University; Mumcu: University of Essex.

We thank ESRC (through MiSoc) and ING Bank (through its Corporate Social Responsibility program) for support. S.E. 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.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00756.

[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 & Vella, 2013; Johnston, Schurer, & Shields, 2014). Olivetti, Patacchini, and Zenou (2016) find that women's work hours are positively affected by both the work behavior of their own mother and their peers' mothers.

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

The Review of Economics and Statistics, December 2018, 100(5): 876–890
© 2018 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology
doi:10.1162/rest_a_00756

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 II.

We collected the data reported in this paper 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. The data set includes a rich set of variables on student, family, and teacher characteristics that we 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 pedagogical approach to teaching, as well as indicators of personal effort, was 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, an effect amplified with longer exposure to the same teacher. If the teacher has been teaching the student for two to three years, a 1 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 the teacher's gender role beliefs that influences girls' beliefs on gender role. 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 & Long, 2005; Dee, 2007; Hoffmann & Oreopoulos, 2009; Carrell, Page, & West, 2010; Antecol, Ozkan, & Ozbeklik, 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. A recent set of papers documents the effects of gender biases as reflected in discrimination in grading on student achievement and choices, with differing results (Lavy & Sand, 2015; Terrier, 2015). That paper finds (in Israel) that boys are overassessed, with negative effects on girls' achievement and future math course choices; it also 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 antimale bias in grading, and Robinson and Lubienski (2011) also find that teachers rate girls more favorably than cognitive scores would suggest. Our paper differs from these other 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 nonparametric 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 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.

II. Background

The Turkish twelve-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 past fifteen years, the majority of the middle- and upper-class parents prefer to send their children to private schools. Our study sample covers third- and fourth-grade students in public elementary schools in particularly needy areas of Istanbul. It therefore primarily represents Turkey's lower

[3] A large literature in economics studies the effect of teacher quality on educational attainment (Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005; Aaronson, Barrow, & Sander, 2007; Harris & Sass, 2011; Hanushek, 2011; Chetty, Friedman, & Rockoff, 2014). See also Hanushek and Rivkin (2006) and Schwerdt & Wuppermann (2011).

socioeconomic segment, with limited variation with respect to socioeconomic status.

In studying the impact of teachers' beliefs on actual student outcomes, one faces a fundamental selection issue: students in a given school may be allocated to teachers in a nonrandom manner. This happens, for example, when a particular type of parent selects a particular type of teacher—one 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 drawings in the presence of parents. Classroom sizes are not allowed to exceed fifty, although a maximum of thirty is typically preferred. Draws are stratified based on gender and preschool attendance to ensure balance in gender and school preparedness in each classroom. Therefore, contrary to the private school system, there is no room in the state system for parents to choose their child's teacher. Of course, parents may decide to send their child to a school that is not in the catchment area; however, acceptance of the student to a noncatchment 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, remixing in later grades is extremely rare, which means that unless the family moves, students remain with the same classmates until they graduate from elementary school.

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 IIIA.

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, reappointments, and retirement, a teacher teaches the students allocated to him or her from grade 1 to grade 4, when they move on to middle school. Because of the strictly centralized allocation of teachers and subsequent reappointments 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.

A. *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 reappointed (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 voice in which city, let alone district or school, she will be appointed to. It is generally very difficult to be appointed to one's preferred city before five to ten years of teaching experience 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 because 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. The only way for a teacher to accumulate service points is by teaching. These points are very important for teachers because they determine their chances of being reappointed to the city of their choice or the district of their choice if they are already in a city they like.

After appointment, a teacher can be reappointed to another school (generally within the same city) if 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) or her reappointment request is honored.[5] A class-

[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 subject to the scrutiny of the ministry in this regard.

[5] Teachers cannot ask to be reappointed before completing at least four years (over six years in actual practice) of service in their current school. Requests to be reappointed are honored if there is a school in need in the preferred district and the teacher has higher service points than his or her

room may lose its teacher because of retirement and resignations, but the most common reason is involuntary rotation due to excess supply and reappointment to another school based on teacher request. When a teacher is reappointed to a new school, she is allocated to a classroom in need of a teacher. Because 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 (because 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 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 reiterate that teacher sorting within a district based on any metric other than service years, which we control for in our regressions, is largely ruled out in this system. In section IV, we show that teachers who have been teaching a class for a longer or shorter time are largely similar in terms of the rich observables we have such as demographics, qualifications, and teaching styles.

III. Data

Our data were collected as part of a large-scale field project underway since 2013. The project aims to study the behaviors, attitudes, and outcomes of students in conjunction with the behaviors and attitudes of teachers. We collected all of the student data 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 our tests and filled in the questionnaires.

Data were collected using a rich battery of tools: 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% nonattendance 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 data set with 31 schools, 145 teachers, and approximately 4,000 students. Our typical teacher is female and university educated, and she has accumulated about fifteen 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 class teaching program, which is a four-year university degree in elementary school teaching.

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

A. *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 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. We prepared and extensively piloted these tests based on national curricula. An independent set of teachers was consulted to tailor the questions to each grade (3 and 4). We measured students' cognitive ability by Raven's Progressive Matrices. We also had access to students' official math, verbal, and behavior grades, all given by their own teachers.

B. *Teacher Characteristics*

The primary purpose of this 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

competitors who have the same location preference. Because 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 like 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.

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

[7] We did not attempt to collect this information directly from parents, as our experience is 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.

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 noncognitive skills) and family background, it has proven to be very difficult to measure.[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 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 of the students, are likely to be important factors in student outcomes (Domino, 1971; Schwerdt & Wuppermann, 2011; Bietenbeck, 2014; Hidalgo-Cabrillana & Lopez-Mayan, 2015). Using item-set questions directed to teachers, we construct four distinct teaching style variables that call modern versus traditional, growth versus fixed mind-set, warm versus distanced, and extrinsic versus intrinsic motivators. A traditional teaching style is reflected in the teacher's dictating to the students what to do in class and following a rigid structure to each class that she determines. What we call a modern approach to teaching involves the students more in the learning process and aims to induce the children to think critically.[9] Having a growth mind-set (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 mind-set has been found in the literature to predict academic achievement (Blackwell, Trzesniewski, & Dweck, 2007; Alan, Boneva, & Ertac, 2016). From the perspective of the teacher, we measure growth mind-set through questions about the relative importance of innate ability versus sustained effort for success (e.g., whether the teacher agrees that any student could become the best in the class by working hard enough). The warm versus distanced construct gets at how authoritarian the teacher is in her interactions with the students and how important it is for 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 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 extracurricular activity by teachers reflects voluntary effort.[11] We therefore collect information on teachers' extracurricular 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 teachers collect service points passively, only by teaching. No other activity or certificate or diploma will matter in collecting the service points required for reappointments, salary increases, and retirement benefits. Nevertheless, many certificate and diploma programs, as well as conferences and social projects, 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 voluntarily, paying participation fees (if any) themselves and sacrificing evening and weekend time. 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.

C. *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 four-point Likert scale, with which we construct a gender stereotype score for each teacher and each student. Some example questions are, "It is more important for boys to go to college than girls," "Women cannot play football well even if they try hard," and "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 have the following answers: "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

[8] See Carrell and West (2010).

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

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

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

FIGURE 1.—DISTRIBUTION OF GENDER ROLE BELIEFS

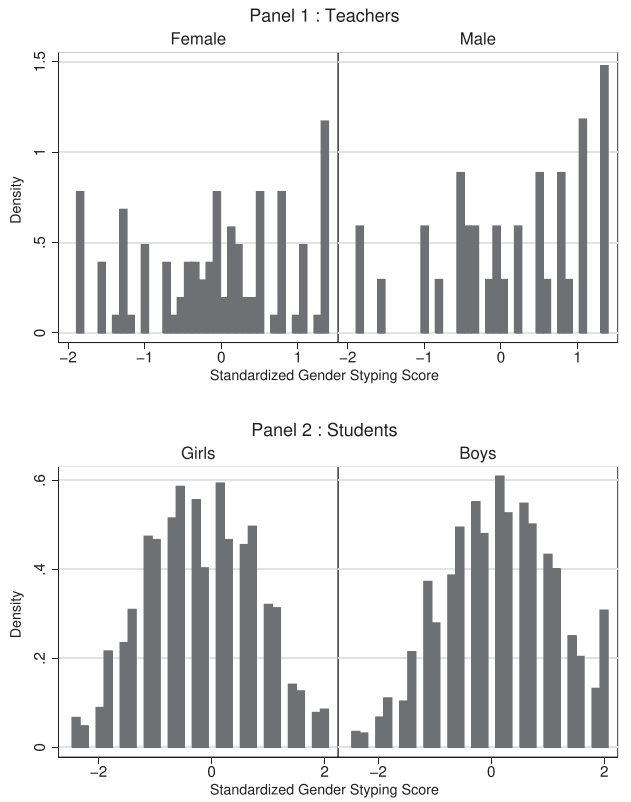


TABLE 1.—PREDICTORS OF TEACHERS' GENDER ROLE BELIEFS

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 mind-set				−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 extracurricular programs					0.005
					(0.01)
Number of volunteer activities					−0.041***
					(0.01)
N	145	145	145	145	145
R^2	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. Significant at $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

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 into it because of their university entrance exam performance) 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^2 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 mind-set 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.

IV. 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 helps us conduct a causal mediation analysis as detailed in section V. 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

[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.

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 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 0 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 last is a question based on a four-item scale that asks how much the father takes part in household chores. Family income or wealth level (SES) is reported by the teacher based on an item scale of 1 to 5, and low-, medium-, and high-SES indicators are constructed based on these.

measure of IQ (elicited via Raven's progressive matrices) is balanced across the two types of teachers.[13]

A. *Empirical Specification*

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

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

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 mind-set, 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.

[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.

We divide the exposure variable into three groups. Children who have been taught by 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 relate 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 gender-differential results presented in section IVB and mediation analysis in section V largely mitigate this concern.

B. *Gender Role Beliefs of Teachers and Student Achievement*

We estimate the empirical model presented in section IVA 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 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 V. Table 3 presents the results by suppressing the coefficient estimates of student, family, and teacher characteristics. Table A.2 in the online appendix, which gives the full results, shows that 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 1 standard deviation increase in the Raven (IQ) score is associated with 0.35 (0.23) standard deviation 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 1 standard deviation increase in the gender stereotype score

[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 in figure A.1.

[15] About 13% of the students have been exposed to the same teacher less than their classmates have. We consider them as relocators. 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 environments, their cognitive and noncognitive skills, including their math and verbal test scores, do not appear to be different from the rest of the sample. Results are available on request.

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	1,870	1,943	1,873	1,946
R^2	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 versus distanced, extrinsic versus intrinsic motivator, traditional versus modern, and growth versus fixed mind-set. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping (gender stereotyping) score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level. Significant at $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

(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 1 standard deviation increase in teachers' gender-stereotyped beliefs lowers girls' test scores in mathematics by about 0.21 standard deviation. The effect for an exposure of two to three years is smaller: a 1 standard deviation increase in teachers' gender-stereotyped beliefs lowers girls' test scores in mathematics by about 0.12 standard deviation. While the equality of coefficients for four-year and two- to three-year exposure is not rejected for either gender, we estimate a statistically significant effect of four-year exposure to the same teacher relative to one-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 deviation for two- to three-year and four-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 four-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. 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.

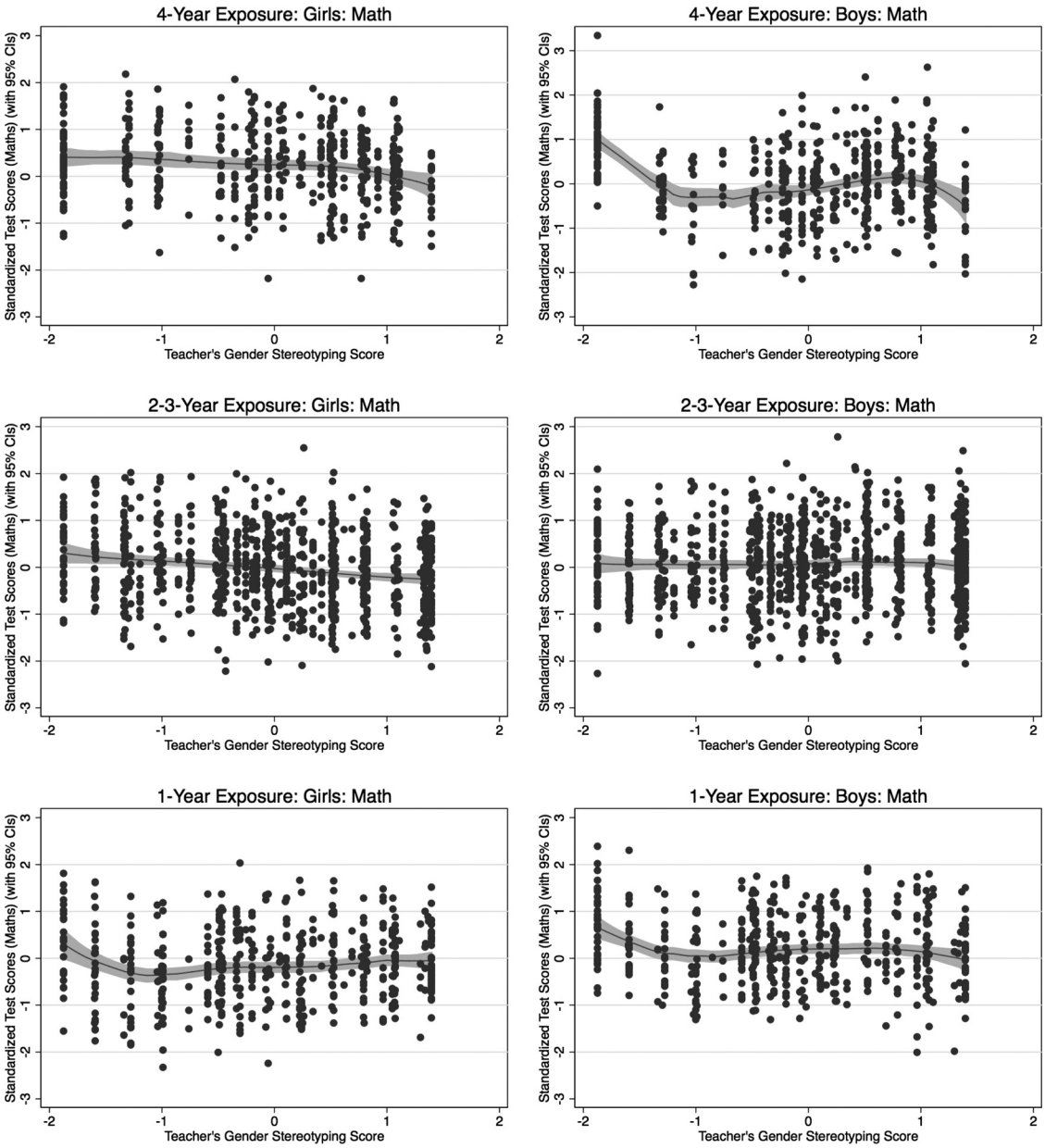
Since our measure of gender role beliefs is a continuous construct, it would be informative to present the functional relationship between test scores and teacher's beliefs in a nonparametric fashion. For this, we relax our assumption of a linear parametric model and modify our empirical model as follows:

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

Here, while all student, family, and teacher characteristics enter the model linearly, we allow for test scores to be a

[16] Related to this result, Hill and Jones (2018) find that repeat student-teacher matches have a significantly positive effect on student achievement in similar (third- to fifth-grade) elementary school students, pointing to the benefit of staying with the same teacher.

FIGURE 2.—TEACHER GENDER STEREOTYPING AND MATH TEST SCORES: NONPARAMETRIC

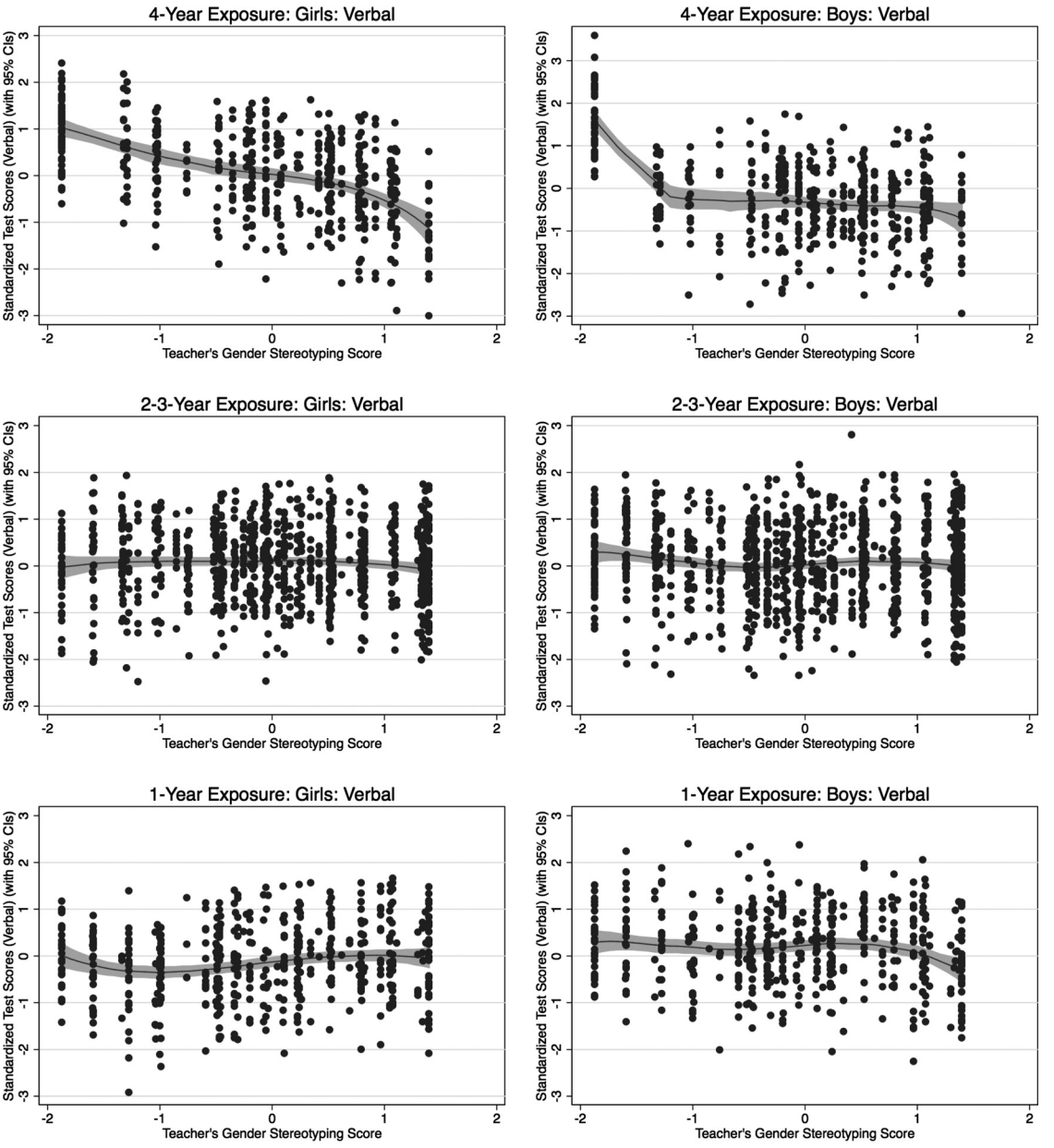


Figures plot the nonparametric estimates (and 95% confidence bands) of the effect of teacher's role beliefs on math test scores for girls (column 1) and for boys (column 2). All student, family, and teacher characteristics enter the model linearly, and school fixed effects are included.

nonparametric 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 reemerge for girls in these pictures. Looking at four-year and two- to three-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 nonlinear 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 one-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

FIGURE 3.—TEACHER GENDER STEREOTYPING AND VERBAL TEST SCORES: NONPARAMETRIC



Figures plot the nonparametric 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.

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 nonparametric specifications, we estimate a declining and fairly linear relationship for girls under four-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.

C. *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 reestimate 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 (fifteen 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.

Although 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

TABLE 4.—HETEROGENEOUS EFFECTS OF TEACHER GENDER ROLE BELIEFS ON GRADES

	Math Score		Verbal Score	
	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	1,594	1,652	1,594	1,652
R^2	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 mind-set, 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 versus distanced, extrinsic versus intrinsic motivator, traditional versus modern, and growth versus fixed mind-set. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. G_Styping (gender stereotyping) score is constructed in a way that larger values indicate more traditional gender role beliefs. Standard errors are clustered at the teacher (classroom) level. Significant at $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

schools. Recall that the institutional structure leaves very little room for self-selection of teachers into catchment areas or schools, and our sample consists of generally "undesirable" schools. However, teachers who have accumulated high service points (those with a higher number of years of service) might be able to self-select into relatively more desirable schools, although this is still difficult. Given that working in a catchment area of one's choice is generally ruled out before twenty years of service except purely by chance, we reestimate our linear model by excluding the teachers who have more than twenty 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 she observes in the class. Suppose that in a given classroom or cohort, boys are indeed better academically than girls. If the teacher bases 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, allow us to address this issue. Our teacher survey includes a question where we ask the teacher whether she has observed boys or girls to be better at math (or equal) in her experience as a teacher. When we exclude the teachers who report boys to be better (only seven teachers), our results remain the same (see table A.5).[17]

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

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. Fifty-six percent of our teachers report that they have observed girls to be better at math, 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—that is, 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.

[18] In our data, the unconditional performance of girls and boys in an objective math test is similar; however, the dummy for males 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 of 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 noncognitive skills and good behavior in addition to pure exam performance, especially in elementary school (Brookhart, 1993; McMillan, Myran, & Workman, 2002; Borghans, Goldsteyn, & Heckman, 2016; Jackson, 2016). Such effects may also potentially explain findings of grading biases in favor of girls (Terrier, 2015).

V. 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 toward 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 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 or focusing on boys more in math. Finally, another potential mediator could be the students' mind-set on achievement—whether students have a growth mind-set that highlights the importance of effort or a fixed mind-set that emphasizes innate abilities. Gender-biased teachers, who hold fixed views of what each gender can and cannot do, may influence the achievement mind-set of students, particularly girls. This shift toward a fixed mind-set may in turn lead to lower motivation and performance, as has been shown in the literature (Blackwell et al., 2007; Alan et al., 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, Davidson, & Nelson, 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, Keele, and Yamamoto (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 states that given the pretreatment confounders, treatment assignment is independent of the potential outcomes and potential mediators. The second states that the mediators are independent of the potential outcomes conditional on pretreatment 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 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, that is, 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 deviation (p-value $= 0.001$). The relationship is not statistically different from 0 for boys (p-value $= 0.96$). We estimate no effect on self-confidence and mind-set 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 deviation. 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

TABLE 5.—MEDIATOR MODEL: THE EFFECT OF TEACHERS' BELIEFS ON STUDENTS' BELIEFS

	Gender Role Beliefs		Self-Confidence		Growth Mind-set	
	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	1,888	1,967	1,888	1,967	1,888	1,967

Dependent variables are standardized scores of students' gender role beliefs, self-confidence, and growth mind-set. The binary variable Traditional Teacher takes the value 1 if the teacher's beliefs are above the median score and 0 otherwise. Student characteristics: student gender, age in months, Raven IQ score, self-reported confidence, gender role beliefs, growth mind-set, 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 versus distanced, extrinsic versus intrinsic motivator, traditional versus modern, and growth versus fixed-mind-set. Teacher effort: Number of voluntary programs for teaching improvement completed and number of voluntary class activities organized for teaching purposes. Standard errors are clustered at the teacher (classroom) level. Significant at ***$p < 0.01$.

TABLE 6.—POTENTIAL CHANNELS FOR THE EFFECTS ON TEST SCORES: CAUSAL MEDIATION

	A. Math Test Scores		
	Gender Role Beliefs	Self-Confidence	Growth Mind-set
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%
	B. Verbal Test Scores		
	Gender Role Beliefs	Self-Confidence	Growth Mind-set
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. 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 1,000. Significant at **5%.

effect of factors such as lower academic attention on girls by traditional teachers.[21]

The results on verbal scores are quite interesting. The total effect of the teacher's gender role beliefs on verbal performance is not statistically different from 0 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 0 for boys.[22]

Note that our analysis shows 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 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 will not 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 drop. In fact, the 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," and "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, provide 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.

[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 on request.

[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.

VI. 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 data set 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 in both 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 nontrivial portion of the effect comes from teachers who transmit their 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, patriarchal gender roles are particularly imposing and improving achievement is a policy imperative (Heckman, 2006).

Two caveats are worth mentioning. First is the fact that our data are 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 & Rufolo, 1990; Joensen & 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

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics* 25 (2007), 95–135.

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

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz, "Personality Psychology and Economics" (pp. 1–181), in E. Hanushek, S. Machin, and L. Woessman, eds., *Handbook of the Economics of Education* (Amsterdom: Elsevier, 2010).

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik, "The Effect of Teacher Gender on Student Achievement in Primary School," *Journal of Labor Economics* 33 (2014), 63–89.

Bertrand, Marianne, "New Perspectives on Gender" (1534–1590), in David Card and Orley Ashenfelter eds., *Handbook of Labor Economics*, vol. 4, part B (Amsterdam: Elsevier, 2011).

Bertrand, Marianne, and Kevin F. Hallock, "The Gender Gap in Top Corporate Jobs," *Industrial and Labor Relations Review* 55 (2001), 3–21.

Bettinger, Eric P., and Bridget Terry Long, "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *American Economic Review* 95 (2005), 152–157.

Bietenbeck, Jan, "Teaching Practices and Cognitive Skills," *Labour Economics* 30 (2014), 143–153.

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

Blackwell, Lisa S., Kali H. Trzesniewski, and Carol Sorich Dweck, "Implicit Theories of Intelligence Predict Achievement across an Adolescent Transition: A Longitudinal Study and an Intervention," *Child Development* 78 (2007), 246–263.

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

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries, "What Grades and Achievement Tests Measure," *Proceedings of the National Academy of Sciences* 113 (2016), 13354–13359.

Brookhart, Susan M., "Teachers' Grading Practices: Meaning and Values," *Journal of Educational Measurement* 30:2 (1993), 123–142.

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *Quarterly Journal of Economics* 125 (2010), 1101–1144.

Carrell, Scott E., and James E. West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy* 118 (2010), 409–432.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review* 104 (2014), 2633–2679.

Dee, Thomas S., "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95 (2005), 158–165.

———— "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources* 42 (2007), 528–554.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger, "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review* 104 (2014), 991–1013.

Domino, George, "Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement," *Journal of Educational Psychology* 62 (1971), 427.

Dweck, Carol S., *Mindset: The New Psychology of Success* (New York: Random House, 2006).
Dweck, Carol S., William Davidson, Sharon Nelson, and Bradley Enna, "Sex Differences in Learned Helplessness: II. The Contingencies of Evaluative Feedback in the Classroom and III: An Experimental Analysis," *Developmental Psychology* 14 (1978), 268.
Farre, Lidia, and Francis Vella, "The Intergenerational Transmission of Gender Role Attitudes and Its Implications for Female Labour Force Participation," *Economica* 80 (2013), 219–247.
Fortin, Nicole M., "Gender Role Attitudes and the Labour-Market Outcomes of Women across OECD Countries," *Oxford Review of Economic Policy* 21 (2005), 416–438.
Fryer, Roland G., "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," *Journal of Labor Economics* 31 (2013), 373–407.
Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales, "Culture, Gender, and Math," *Science* 320:5880 (2008), 1164–1165.
Hanushek, Eric A., "The Economic Value of Higher Teacher Quality," *Economics of Education Review* 30 (2011), 466–479.
Hanushek, Eric, and Steven Rivkin, "Teacher Quality" (pp. 1051–1078), in Eric Hanushek and Finis Wekh, eds., *Handbook of the Economics of Education* vol. 2 (Amsterdam: North-Holland, 2006).
Harris, Douglas N., and Tim R. Sass, "Teacher Training, Teacher Quality and Student Achievement," *Journal of Public Economics* 95 (2011), 798–812.
Heckman, James J., "Skill Formation and the Economics of Investing in Disadvantaged Children," *Science* 312:5782 (2006), 1900–1902.
Hidalgo Cabrillana, Ana, and Cristina Lopez-Mayan, "Teaching Styles and Achievement: Student and Teacher Perspectives," Universidad Autónoma de Madrid working papers in economic theory (2015).
Hill, Andrew J., and Daniel B. Jones, "A Teacher Who Knows Me: The Academic Benefits of Repeat Student-Teacher Matches," *Economics of Education Review* 64 (2018), 1–12.
Hoffmann, Florian, and Philip Oreopoulos, "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources* 44 (2009), 479–494.
Imai, Kosuke, Luke Keele, and Teppei Yamamoto, "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science* 25:1 (2010), 51–71.
Jackson, C. Kirabo, "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes," NBER working paper 22226 (2016).
Joensen, Juanna Schrøter, and Helena Skyt Nielsen, "Is There a Causal Effect of High School Math on Labor Market Outcomes?" *Journal of Human Resources* 44 (2000), 171–198.
Johnston, David W., Stefanie Schurer, and Michael A. Shields, "Maternal Gender Role Attitudes, Human Capital Investment, and Labour Supply of Sons and Daughters," *Oxford Economic Papers* 66 (2014), 631–659.
Lavy, Victor, "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment," *Journal of Public Economics* 92 (2008), 2083–2105.
Lavy, Victor, and Edith Sand, "On the Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases," NBER working paper 20909 (2015).
McMillan, James H., Steve Myran, and Daryl, Workman, "Elementary Teachers' Classroom Assessment and Grading Practices," *Journal of Educational Research* 95 (2002), 203–213.
Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou, "Mothers, Peers and Gender Identity," Boston College working paper 904 (2016).
OECD, *TALIS User Guide for the International Database* (2013). http://www.oecd.org/edu/school/TALIS-2013-User-guide.pdf.
Paglin, Morton, and Anthony M. Rufolo, "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences," *Journal of Labor Economics* 8:1, pt1 (1990), 123–144.
Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica* 73 (2005), 417–458.
Robinson, Joseph Paul, and Sarah Theule Lubienski, "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 (2011), 268–302.
Rockoff, Jonah E., "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review* 94 (2004), 247–252.
Sadker, Myra, and David Sadker, *Failing at Fairness: How America's Schools Cheat Girls* (New York: Simon and Schuster, 2010).
Schwerdt, Guido, and Amelie C. Wuppermann, "Is Traditional Teaching Really All That Bad? A Within-Student Between-Subject Approach," *Economics of Education Review* 30 (2011), 365–379.
Spencer, Steven J., Claude M. Steele, and Diane M. Quinn, "Stereotype Threat and Women's Math Performance," *Journal of Experimental Social Psychology* 35 (1999), 4–28.
Terrier, Camille, "Giving a Little Help to Girls? Evidence on Grade Discrimination and Its Effect on Students' Achievement," CEP discussion paper 1341 (2015).