

# Role Models, the Formation of Beliefs, and Girls' Math Ability: Evidence from Random Assignment of Students in Chinese Middle Schools

Alex Eble and Feng Hu\*

February 2017

## Abstract

This paper studies the power of role models in changing beliefs, behavior, and outcomes for children who are struggling in school and who face negative stereotypes. We exploit random assignment of Chinese middle school students to classrooms to estimate how being assigned a female math teacher affects low performing girls. In our data, there is widespread belief that men are more able to learn mathematics than women. We find that teacher-student gender match improves math test scores by 0.45 SD for low performing girls, increases their beliefs in the ability of women to learn mathematics, reduces their perception of the difficulty of learning math, and increases their investment in math-related human capital. These results are consistent with the main predictions of a model of aspirations and aspirations failure. They also provide direct empirical evidence for a specific mechanism, the power of role models in shaping beliefs, driving the common finding that a same-gendered teacher improves girls' performance.

---

\*Eble: Teachers College, Columbia University. Email: eble@tc.columbia.edu Hu: School of Economics and Management, University of Science and Technology Beijing feng3hu@gmail.com. The authors are grateful to John Friedman, Asim Khwaja, and Ilyana Kuziemko for generous feedback. Key words: gender; belief formation; stereotypes; human capital; cognitive skills; behavioral economics. JEL codes: I20; J16; O15

# 1 Introduction

Negative gender norms begin to form early in life. Between the ages of 5 and 7, both boys and girls start to perceive women as less likely to be of high ability, and this affects the interests of both sets of children (Bian et al., 2017). These perceptions lead to worse performance on tests (Spencer et al., 1999), which has been shown to negatively impact later life outcomes (Lavy et al., 2014). While we know that being paired with a teacher of the same gender can improve girls' performance (Dee, 2007; Muralidharan and Sheth, 2016), particularly in fields where stereotypes against women persist (Carrell et al., 2010), we have little direct evidence about the mechanisms driving this effect. This paper provides positive empirical evidence from China in favor of one hypothesized channel: the power of role models to shape beliefs and investment behavior.

Students make forward-looking human capital investment decisions with limited information, and these decisions are likely to be influenced by the informational environment around them. The lack of a positive, credible female role model in relevant areas (e.g., STEM fields) for girls to aspire to could lead students to draw incorrect inference on the returns to schooling, i.e., if they lack examples of individuals with high returns they will assume such returns do not exist. This, in turn, would lead to negative stereotypical beliefs and suboptimal investment behavior. The sociological work of William Julius Wilson has hypothesized that this lack of positive role models is one reason for underinvestment in human capital in inner city America (Wilson, 2012). The model of "aspirations failure" laid out in Genicot and Ray (Forthcoming) predicts that this risk is particularly high for students with lower endowments of skill, family privilege, or other characteristics which facilitate success in school, as these factors also drive down the ratio of perceived benefits of investment in human capital to the costs.

In this paper, we provide evidence from Chinese middle schools to answer two questions related to this proposed mechanism: one, does being taught by a female teacher change beliefs about own mathematical ability, and that of their gender, for girls who are struggling in school? Two, how do gender beliefs manifest in behavior (investment in human capital) and performance (test scores) among this group? Using random assignment of students to classrooms, we estimate the effect of having a female math teacher on the beliefs, efforts, and academic performance of girls in the left tail of the ability distribution. We find that both for girls who have fared poorly in math in primary school, and for girls to the left of the median middle school math test score, being assigned a female

teacher changes test scores, gender stereotypes, perceived difficulty of math, and effort.

To conduct this analysis, we use the baseline wave of the China Education Panel Survey (CEPS), a nationally representative survey of Chinese middle school students. The survey collected information from school administrators on whether students are randomly assigned to classrooms or are assigned to classrooms through non-random mechanisms (mostly tracking). The CEPS also elicited detailed information on students' gender-specific stereotypical beliefs regarding math learning ability, on student-teacher interactions in the classroom, and on student time use. These data make it possible to explicitly investigate underlying mechanisms probed only indirectly in other work on the impact of teacher-student gender match on girls (Antecol et al., 2015; Carrell et al., 2010; Dee, 2007; Paredes, 2014).

We find that being assigned a female math teacher generates a 0.45 SD improvement in the math test scores of low performing girls. This assignment also changes beliefs and behavior. Low performing girls assigned to a female math teacher are 17 percentage points more likely to disagree with the statement that boys are better at learning math than girls (from a baseline of 45%), are nearly 20 percentage points less likely to find math extremely difficult (baseline 80%), and are 10 percentage points more likely to enroll in mathematics tutoring (baseline 15%). We find no evidence that female math teachers favor girls in class with more praise or attention, particularly not the low performing girls for whom we see the largest change in beliefs, effort, and performance. Finally, we show that the gender-specific benefits we observe are unlikely to be associated with any difference in teacher quality between female and male teachers.

This paper contributes to two literatures. First, we further the budding literature on aspirations and aspirations failure by making explicit an empirical link between role models, the formation of beliefs, and behavior that can affect long-term outcomes. Our results are consistent with both the hypothesis of Wilson (2012) and a key prediction of the model in Genicot and Ray (Forthcoming); namely, that presence of a plausible role model may induce changes in beliefs about one's own chances in the world, behavior, and outcomes. Second, we contribute to the longstanding literature on the effects of teacher-student gender match (Dee, 2007; Carrell et al., 2010; Muralidharan and Sheth, 2016). While this literature has hypothesized and shown indirect evidence for possible mechanisms driving the largely positive effects found (Paredes, 2014), we provide the first direct evidence we are aware of in support of a specific mechanism, the power of role models to shape beliefs, in driving the positive effects of teacher-student gender match on student test scores.

The rest of this paper is structured as follows. In Section 2 we conduct an empirical exercise and lay out a conceptual framework to motivate the focus of the paper on low performing girls. Section 3 describes the setting we study. Section 4.1 outlines our data sources and provides summary statistics of our main variables and Section 4.2 introduces our empirical strategy and presents results for tests of our main identifying assumptions. Section 5 presents our main results, Section 6 evaluates evidence for each in a set of possible underlying mechanisms, and the final section concludes.

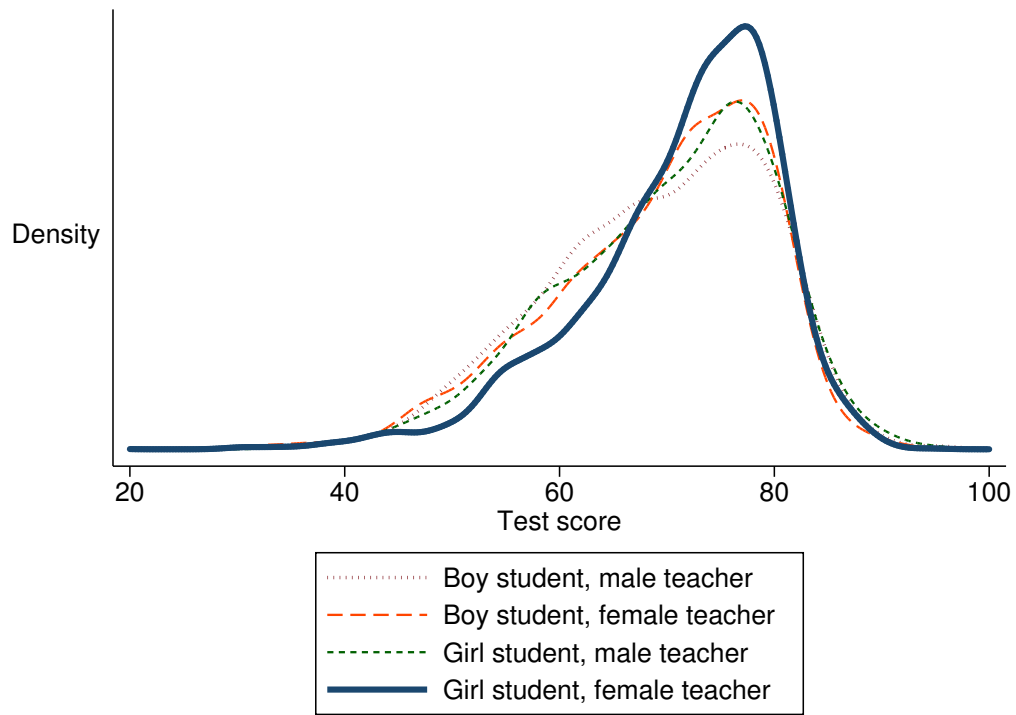
## **2 Conceptual framework and empirical motivation**

This section motivates our empirical analysis. First, we discuss the empirical evidence for the our focus on “low performing” girls, and we then we outline a simple conceptual framework, drawing on Genicot and Ray, which generates predictions that we test later in the paper.

We first examine the empirical distribution of math test scores in our data separately by teacher-student gender configuration. In Figure 1, we show a kernel density plot of math test scores for the four different teacher-student gender pairings (FF, MF, FM, and MM). There is a substantial gain in the left tail of the distribution for girls assigned a female math teacher relative to all other pairing types. A Kolmogorov-Smirnov test rejects the equality of the FF distribution from the combined distribution of the test scores of students in other teacher-student gender pairings with a p-value of less than 0.001, and quantile regressions show substantial gains in the first through sixth deciles. These results suggest that the we should look among these girls, i.e., low performing girls, for other potential impacts of teacher-student gender match.

Next, we place our analysis in the context of relevant literature in economics and psychology and derive predictions from an elementary version of the model of aspirations and aspirations failure laid out in Genicot and Ray (Forthcoming). Both across countries and in our Chinese data, girls express a lack of confidence in their own abilities in math and the math ability of their gender (Beilock et al., 2010; OECD, 2015). The empirical literature in psychology demonstrates that this type of “gender-stereotyping” belief in girls may either contribute to worse performance directly, through anxiety because of “stereotype threat” (Cheryan, 2012; Niederle and Vesterlund, 2010; Shih et al., 1999; Spencer et al., 1999), or could cause girls to invest less time in studying for math, relative to other subjects, thus generating a self-fulfilling prophecy (Bian et al., 2017).

Figure 1: Distribution of math test scores by teacher-student gender pairing



Notes: This figure plots the distribution of students' scores on math midterm examinations by teacher-student gender configuration. The sample is restricted to the estimation sample used in the previous tables. A gaussian kernel was used to generate the density plots. A Kolmogorov-Smirnov test rejects equality of the distributions of test scores between two groups: girls paired with a female teacher and the combined distribution of students in all other teacher-student gender configurations. Test scores are standardized within grades and schools so that ten points is one standard deviation and the mean is 70.

One corollary of these findings is that the presence of a positive female role model could change girls' views about the potential positive returns to their effort in math by providing an example of a plausible (by virtue of shared gender) example of the returns to such effort (Carrell et al., 2010; Genicot and Ray, Forthcoming; Wilson, 2012). This, in turn, could change girls' willingness to exert effort in the subject area (Beaman et al., 2009; Gunderson et al., 2012; Nixon and Robinson, 1999). Evidence from psychology also suggests that such an example could lead to an increase in students' academic motivation and expectations (Nixon and Robinson, 1999).

The Genicot and Ray model posits that individuals choose an amount of productive investment in themselves (e.g., by choosing how much to study, whether to pursue further education, or whether to invest in training) based on the distance they see between their current state and an aspired-to future, henceforth simply their "aspirational distance." This distance is a function of the perceived amount of personal input needed to reach the desired outcome, which in turn is a function of one's own endowment and the informational environment around the individual. Greater distance is a function of both a perception of greater input needed to reach one's aspirations and greater uncertainty about the possibility of reaching them at all.

Genicot and Ray show that there is an inverted-U relationship between the aspirational distance and the choice of investment. At very low levels of aspirational distance, investment will be low, because there is little investment needed for the individual to reach her aspirations. As distance increases, investment increases, as the aspiration is still attainable but it requires more input. Beyond a certain point, however, investment again decreases, as the distance becomes great enough that the perceived benefit of investment net of costs decreases in part as a result of increased uncertainty of the outcome.

Assuming a common aspiration across all children (i.e., success in math), a key prediction of the model is as follows: because of low ability levels and societal pressures, girls who are lower in the ability distribution have higher aspirational distance, as their personal characteristics (signals about their own ability from test scores) and environmental influences (negative gender stereotypes) lead them to believe that either the necessary amount of investment for them to reach their aspirations or the uncertainty surrounding the returns to this investment may be too large to justify. In our data, we see evidence of negative stereotypical beliefs consistent with large aspirational distance: fifty two percent of our sample (boys and girls) believes that boys are better than girls at learning math. Girls to the left of the median math test score are 15 percentage points more likely than girls to the

right of the median to believe that boys are better at math than girls (54 percent vs. 39 percent). Furthermore, despite performing better on most math tests, girls are more likely to report that they find math difficult.

This generates a prediction about the effect of teacher-student gender match on beliefs and effort/investment that we will take to the data. Specifically, we argue that being assigned a female math teacher provides low performing girls with a plausible example of success in math which may reduce uncertainty about the possible positive returns to investment in math-related human capital. The model predicts that this should reduce aspirational distance and, in so doing, change beliefs and increase investment. For boys, who have no negative stereotypical beliefs about their ability, and for girls who are doing well in math, we should see no change in beliefs or investment, as these groups have more evidence of their own ability, and so a positive example is less likely to affect aspirational distance.

### **3 Key details on Chinese middle school education**

China's 1986 compulsory education law mandated that all children receive nine years of free compulsory education, including six years of primary schooling (the first to sixth grades) and three years of middle school education (seventh to ninth grade). Until the late 1990s, primary school graduates were required to attend an entrance examination to be eligible to enter middle school (Carman and Zhang, 2012; Lai et al., 2011). At the turn of the millennium, middle schools were prohibited from admitting students based on academic merit and the middle school entrance examination was later cancelled. In the same spirit, tracking of students to different classes based on demonstrated ability or academic performance has been banned in middle school since the latest compulsory education law was issued in 2006.

There are two permitted methods of assigning students to classes in China's middle schools: (1) purely random assignment and (2) assignment of students to maintain similar average levels of performance across classes, based either on students' academic performance on primary school graduation examinations or on diagnostic examinations arranged by the middle school. In the first system, primary school graduates are assigned to a neighborhood middle school according to the needs of local educational authorities, and then they are randomly assigned to classes by a lottery

or another quasi-random method<sup>1</sup>. In the second system, students are assigned to classes by an algorithm which takes into account their academic performance at the beginning of the seventh grade and enforces a “balanced assignment” rule. This rule requires that the average quality of students be comparable across classes (Carman and Zhang, 2012).

To understand this second rule, consider the following example. Assume that one middle school has a total of 200 incoming seventh-grade students, who will be assigned to five classes. Students are first ranked by their total scores on primary school graduation examinations and then are assigned to classes according to their score ranks. For example, the top five and the bottom five students are assigned to each of the five classes respectively, with the best (ranked first) and the worst (ranked 200th) students in the same class. That is, the average rank of students in each class –  $(1+200)/2$  for class one,  $(2+199)/2$  for class two,  $(3+198)/2$  for class three, and so on — is kept about the same, in this case, 201/2.

This system is not implemented with perfect fidelity, however, particularly as students move beyond the first grade of middle school (i.e., from the seventh grade to the eighth). Unlike the system in many western countries, where admission to high school or university is based on multiple dimensions such as teacher recommendations and personal leadership potential, China’s secondary school admissions system relies almost exclusively on entrance examinations (Zhang, 2016). Furthermore, the promotion of middle school administrators (like government officials) is largely determined by the school’s performance in the high school entrance examination, that is, according to the annual number of graduates admitted to elite high schools. To prepare for the entrance examination, therefore, some middle schools assign students to classes based on their academic performance despite the banning of class tracking according to the compulsory education law. Accordingly, school administrators channel more resources (for instance, high-quality teachers) to classes with high-ability students so that they have a better performance record at the high school entrance examinations. This means that after their first semester or year of middle school, students may be reassigned to different classes based on their academic performance even if they are randomly assigned at the beginning of the seventh grade. In this analysis, we restrict our attention to students assigned to classes randomly in the 7th grade and in those schools where random assignment of students to classes is maintained throughout middle school.

---

<sup>1</sup>For instance, according to alphabetical order by surname.



## **4 Data and empirical strategy**

This section describes our data sources and empirical approach. Section 4.1 outlines the data we use and provides summary statistics. Section 4.2 describes the identification strategy we use, stating and testing our identifying assumptions.

### **4.1 Data sources**

The main data source for this paper is the baseline wave of the China Education Panel Survey (CEPS) conducted by the National Survey Research Center at Renmin University of China. The CEPS is a nationally representative longitudinal survey that aims to track middle school students through their educational progress and later labor market activities throughout their life cycles. The baseline survey of the CEPS adopted a stratified, multistage sampling design with probability proportional to size, randomly selecting approximately 20,000 seventh and ninth grade students from 438 classes in 112 schools from 28 counties across mainland China during the 2013-2014 academic year. For each selected school, two classes were randomly chosen for both the seventh and ninth grades, and then all students in the selected classes were surveyed. The CEPS uses five different questionnaires, administered to students, parents, homeroom (banzhuren) teachers, main subject (math, Chinese, and English) teachers, and school administrators, respectively. It is China's first nationally representative survey targeting middle school students, which is comparable to similar surveys in developed countries such as the Adolescent Health Longitudinal Studies (AddHealth) in the U.S. and the National Education Panel Survey (NEPS) in Europe.

The CEPS contains detailed information on students' academic performance, the first outcome of interest in this paper. It collects administrative school records on students' midterm test scores in the following three compulsory subjects: math, Chinese, and English. The scores are standardized in terms of school and grade, with a mean of 70 and a standard deviation of 10. In addition to students' individual characteristics, the CEPS also collects family background information such as number of siblings, parents' education, and household income levels, all potential determinants of students' academic performance.

The CEPS teacher questionnaire contains rich information on teacher characteristics, including teachers' age, gender, education levels, years of teaching experience, whether the teacher graduated from a university for teachers, whether the teacher holds a senior professional rank, and

whether the teacher has won any teaching awards at various levels. The survey also contains information on the subject and the class the teacher taught during the 2013-2014 academic year. We limit most of our analyses to the matched math teacher-student dataset.

Our survey collects data on whether students were assigned to classrooms or not, i.e. whether assignment was “sui ji”, the literal translation of which is “by machine”, i.e., random, according to the average-equilibrating algorithm, or through other methods. In our identification strategy, we will treat the first two assignment mechanisms as good as random, in order to causally estimate the effect of teacher gender on students’ academic performance. About 85% of middle schools assigned entering students to classes in either a random or an average-equalizing manner. Among those schools, one third reassigned students based on past academic performance when they entered the eighth or ninth grade. In our analysis, we will treat assignment to class as random for seventh graders in those schools reporting either randomly assigning or using the average-equalization algorithm to assign seventh-grade students to classes, and for ninth graders in the subset of these schools which also report not reassigning eighth and ninth grade students to new classes in terms of previous academic performance.

Appendix Table A.1 presents summary statistics for students by gender for those students assigned randomly to classrooms. The average age of girls is younger than for boys, and girls are more likely to have more educated parents and higher family incomes. Girls in our sample also have more siblings than boys, a consequence of the prevailing son-favoring tradition and the birth control policy in China, which allows for multiple children in some cases if the first child is a girl. Finally, girls perform better than boys on math tests administered in class.

Table A.2 shows summary statistics for teachers in the classrooms studied in Table A.1. In our data, 39% of the students are taught by male math teachers, alleviating the challenge faced in Antecol et al. (2015), where there was an insufficient number of male teachers. Female math teachers are on average younger and less experienced than their male counterparts. However, female teachers appear to be more qualified than their male counterparts in terms of education, likelihood of holding a senior professional rank, and proportion having won a teaching award at the province or national level<sup>2</sup>.

The significant differences in characteristics between girls and boys and between female and

---

<sup>2</sup>A teaching award at the national level is the most prestigious, followed by an award at the province level, and awards at the city level are the least prestigious.

male math teachers above may reflect certain gender-specific patterns at the region or school level. For instance, girls and female teachers may be more likely to come from urban schools. In the empirical analysis, we investigate potential sources of bias stemming from such unobservable heterogeneity. As we discuss in the next subsection, our empirical strategy compares male and female teachers within a grade within a school. We show that these observed differences attenuate dramatically and cease to be significant at this level of comparison.

## **4.2 Empirical strategy**

In this subsection we outline our empirical strategy. We first discuss our approach to estimating the effects of being assigned a female teacher on female students and on male students. We then test the identifying assumptions we must satisfy in order to interpret our coefficient estimates causally.

Most prior work on the effects of the teacher-student gender match faces the challenge of how to causally estimate the effect of teacher gender on student achievement in the face of nonrandom sorting of students to classrooms or to teachers. Decision rules for assigning teachers to students could generate either upward or downward bias on estimates of the effect we are after. For example, a system of tracking which assigned more successful students to more qualified teachers (which are more often than not women, as we saw in the previous section) could bias results upward. Conversely, efforts to compensate poor student performance with more able teachers could generate a downward bias on these estimates.

Most work to date in this vein addresses this endogeneity concern by controlling for students' prior performance, either through controlling for potential confounders or using fixed effect estimates with panel data (Ehrenberg et al., 1995; Muralidharan and Sheth, 2016; Paredes, 2014). Rothstein (2010), however, argues that estimates generated this way may still be biased. Other studies (Dee, 2007; Holmlund and Sund, 2008) use within-student variation in teacher gender match to address the non-random sorting concern, exploiting teacher turnover or student interactions with different subject teachers after conditioning on student fixed effects. This approach may also fail if the assignment of students and teachers is jointly determined with gender-specific unobservable characteristics (Dee, 2007; Lim and Meer, Forthcoming). Several of these papers acknowledge that randomization of students to teachers would improve the reliability of the results presented (Antecol et al., 2015; Dee, 2007; Lim and Meer, Forthcoming).

In this paper we have such desired conditions, and exploit them to estimate the impact of teacher-student gender match on girl students' performance in mathematics and their beliefs regarding gender-specific math ability. To generate our estimates, we use a reduced form regression, controlling for grade-by-school fixed effects and a vector of observable, predetermined characteristics at the child and teacher levels. Specifically, to determine whether teacher gender differentially affects the outcomes of interest for boys and girls, we estimate the following equation using CEPS data:

$$Y_{icgj} = \beta_0 + \beta_1 G_{icgj}^s + \beta_2 G_{cgj}^t + \beta_3 (G_{cgj}^t * G_{icgj}^s) + \beta_4 X_{icgj} + \beta_5 Y_{cgj} + \eta_{gj} + \epsilon_{icgj} \quad (1)$$

The variables are defined as follows.  $Y_{icgj}$  denotes the outcome of interest. For our main analysis, this will be the midterm math test score for student  $i$  in class  $c$  of grade  $g$  in school  $j$ .  $G_{icgj}^s$  is an indicator equal to one if student  $i$  is female, and  $G_{cgj}^t$  is also an indicator, equal to one if the teacher in class  $c$  is female.  $X_{icgj}$  is a vector of predetermined characteristics at the student level,  $Y_{cgj}$  is a similar vector for teachers,  $\eta_{gj}$  is a grade-by-school fixed effect (included in the model to account for the fact that students are randomly assigned within the same grade in a school, and because we have students from both the 7th and 9th grades in our sample), and  $\epsilon_{icgj}$  is a robust standard error, clustered at the school level to allow for heteroskedasticity and arbitrary serial correlation across students within each school. Unless otherwise specified, the controlled-for student-level characteristics determined prior to assignment of the teacher gender include age, ethnicity (either Han or non-Han), hukou status (agricultural or not), parents' education levels, the child's number of siblings, and a categorical measure of household income. The teacher-level predetermined characteristics include age, education level, years of work experience, whether the teacher graduated from a normal (i.e., teacher training) university, whether the teacher holds a senior rank, and whether she or he has won teaching awards at city, province, or national level.

Intuitively, this strategy compares the academic performance of students who study in the same grade in a middle school and share background characteristics, but are randomly assigned to either a female or male math teacher. Our identifying assumption is that  $G_{icgs}^s$  and  $G_{icgs}^t$  are orthogonal by virtue of random assignment. We test this assumption by examining evidence of and testing for balance on observable characteristics.

There are two parameters of central interest in this paper. The first is  $\beta_3$ , which we interpret

as a quasi-experimental estimate of the benefit to girls of having a female math teacher relative to the benefit to boys of having a female math teacher. The second parameter is the sum  $\beta_2 + \beta_3$ , which captures the total effect on girls of being paired with a female teacher. If our assumption of orthogonality is satisfied, estimating Equation 1 using OLS should recover unbiased estimates of these parameters. As suggested by our conceptual framework, we should also estimate this equation separately for students based on their place in the ability distribution.

We test our hypothesis that within a grade within a given school, the gender of the teacher is randomly assigned by testing that teacher gender is uncorrelated with observable teacher and student-level characteristics that could influence student performance. These characteristics include students' gender, age, ethnic minority status, and agricultural hukou status, parents' education levels, number of siblings, and household income level. We display these estimates in Table 1.

Our estimates in the first column suggest that the unconditional probability of having a female math teacher is strongly correlated with most background variables. Students with more material and social advantage are more likely to be taught by female math teachers. This advantage persists on several levels - children with female teachers are more likely to hold an urban hukou, and have fewer siblings, more educated parents, and richer families. This variation, however, comprises both within- and between-school differences in these characteristics. In the second and third columns, we make the same comparison after netting out school and grade-by-school fixed effects, respectively. After this transformation of the data (i.e., netting out fixed effects), the correlations are an order of magnitude smaller and are statistically insignificant in all but one case. We conclude from this analysis that students' observable predetermined background characteristics are balanced along the gender of math teachers within the same grade in a given school.

This evidence supports our main identifying assumption, but we cannot rule out the possibility that in some cases influential parents or individuals successfully lobbied to be placed with a better teacher. We conclude from the lack of statistically significant correlation between the gender of the teacher and the characteristics of the children shown in columns 2 and 3 of Table 1 that such non-random matching of teachers to children is unlikely to be common enough to substantially bias our estimates. Nonetheless, this could exert an upward bias on the estimates we generate relative to what they would be in a context with perfect fidelity of implementation. As we rely on teachers' reports of whether they use tracking or random assignment, it may also be the case that some

schools who report using random assignment in fact use tracking. In either case, we would expect this to bias upward our estimates of the effect of female teachers on the best students. We show in the next section no significant effects of teacher-student gender match on high performing girls.

Another descriptive comparison of interest is teacher quality across genders. We are interested in investigating the effect of female teachers on student achievement. To do so, we need to assess whether male and female teachers differ on observable characteristics which could drive any effects we measure (Antecol et al., 2015; Cho, 2012). Table 2 reports the estimation results for regressing the following teacher quality indicators on teacher gender: a dummy for having a full-time bachelor degree or above, a dummy for having attended a normal university, years of teaching experience, a dummy for having a senior professional rank, and two dummies for winning teaching awards at different levels. After conditioning on grade-by-school fixed effects, there are no sizable differences in the aspects of teacher quality we are able to observe between female and male math teachers.

## **5 Teacher-student gender match and performance on math tests**

In this section we examine the effect of teacher-student gender match on student ability based on different measures of the student's ability using the empirical approach outlined in the previous section, quantifying the differences apparent in Figure 1. We estimate the effect of teacher-student gender, first for the entire sample, and then separately for students who experienced substantial, some, or no difficulty in mathematics in primary school, as this characteristic was determined prior to the random assignment of children to teachers. We define those girls who report having much difficulty in learning math in the sixth (and final) grade of primary school as "low performing." Average students are those who report having some difficulty, and the high performers are those who report math in primary school being not difficult. The results are presented in Table 3. The first column reports results for estimating Equation 1 using the entire sample, and the second, third, and fourth column report results generated by estimating the equation using only the low, average, and high performers, respectively. Our main focus in this table and subsequent tables will be the results presented in column 2, that is, for the low performers.

We find that having a female math teacher instead of a male one increases the math test scores of low performing girls relative to low performing boys by 4.5 points, or 0.45 sample standard deviations, controlling for other characteristics as in Equation 1. Girls who face no academic disad-

Table 1: Test of balance between students with female and male teachers

	(1)	(2)	(3)
<b>Panel A: Individual Characteristics</b>			
Female	0.006 (0.014)	0.001 (0.016)	0.008 (0.013)
Age	-0.392** (0.191)	-0.228 (0.23)	-0.066** (0.033)
Minority	-0.108 (0.071)	-0.004 (0.009)	0.009 (0.010)
Agricultural hukou	-0.129*** (0.045)	-0.037 (0.026)	-0.03 (0.031)
<b>Panel B: Household Characteristics</b>			
Father's years of education	1.005*** (0.341)	0.16 (0.136)	0.182 (0.164)
Father: high school or more	0.125*** (0.047)	0.021 (0.021)	0.033 (0.026)
Mother's years of education	1.514*** (0.5)	0.177 (0.146)	0.22 (0.184)
Mother: high school or more	0.128*** (0.045)	0.016 (0.019)	0.019 (0.02)
Number of siblings	-0.238*** (0.087)	-0.021 (0.033)	-0.053 (0.043)
Family is "poor"	-0.094** (0.037)	-0.004 (0.019)	0.000 (0.021)
School fixed effects		X	
Grade by school fixed effects			X

Notes: This table shows results from separate regressions of student and household characteristics, listed in the first column, on teacher gender. Robust standard errors clustered at the school level are shown in parentheses. There are 8,345 observations used in this analysis, comprising all those in the matched math teacher-student dataset with non-missing observations for the main dependent and listed independent variables. Observations are at the student level. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

Table 2: Tests for gender-specific teacher quality

	(1)	(2)	(3)
Full-time bachelor's degree or above	0.166** (0.071)	0.01 (0.095)	0.032 (0.131)
Attended a normal university	-0.054* (0.03)	-0.053 (0.034)	-0.065 (0.05)
Years of teaching experience	-2.808** (1.238)	-0.91 (1.739)	0.581 (2.762)
Holds a senior professional rank	0.019 (0.061)	0.04 (0.082)	-0.065 (0.114)
<i>Received teaching award</i>			
at the province or national level	0.025 (0.049)	0.09 (0.077)	0.065 (0.114)
at the city level	-0.037 (0.071)	0.017 (0.100)	-0.032 (0.131)
School fixed effects		X	
Grade-by-school fixed effects			X

Notes: This table shows results for six separately estimated regressions per column, regressing the outcome listed in the first column on the teacher's gender. The 207 observations used to generate this table are at the teacher level.



Table 3: Effects on student math score, by primary school performance in math

	All (1)	Low (2)	Average (3)	High (4)
Female student ( $\beta_1$ )	0.678 (0.566)	1.261 (1.467)	2.386*** (0.611)	1.592** (0.689)
Female teacher ( $\beta_2$ )	1.553** (0.736)	3.459* (2.034)	2.574*** (0.848)	0.925 (0.715)
Female student * Female teacher ( $\beta_3$ )	0.931 (0.626)	4.503*** (1.891)	-0.118 (0.683)	-0.060 (0.746)
$\beta_2 + \beta_3$	2.484***	7.962***	2.456***	0.865
Observations	8,345	850	4,513	2,931

Notes: The dependent variable is the student's math test score. Robust standard errors clustered at the school level are shown in parentheses, and the analysis is estimated using the specification in Equation 1. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

vantage, in contrast, appear to gain no additional (gender-specific) benefit from being assigned a female teacher. Unlike the many previous studies on teacher-student gender match (Antecol et al., 2015; Carrell et al., 2010; Dee, 2004; Paredes, 2014), we study a setting where, overall, girls perform better than boys. Nonetheless, negative gender norms about girls' math ability persist, and are strongest among the low performers. It is precisely among these girls that we see the largest impact of having a female math teacher on math test scores.

In Appendix Table A.4, we show similar results, only defining low performers instead as those whose math test score is below the median value of his or her teacher-gender pairing group (e.g., the median score of the boys paired with a male teacher). This table shows largely similar results - the effects of teacher-student gender match on girls' math test scores are positive and statistically significant only for girls who are paired with male teachers. While these results are substantially smaller than our estimates generated using performance in primary school, (around 0.12 SDs here, as opposed to the 0.45 SD change we measure for the poor performers in primary school), they are nonetheless both sizable and consistent with the predictions of the model. Furthermore, quantile results find very similar effects (more than 0.4 SDs) for students in the 20th quantile. Taken together,

the empirical results presented in this section suggest that having a female math teacher instead of a male one yields a larger benefit to girls than to boys, and that these benefits are concentrated among low performing girls.

## **6 Testing predictions and alternative explanations**

In this section we examine two potential mechanisms which could drive the results presented in the previous section. The first mechanism is that predicted by the Genicot and Ray (Forthcoming) model: that students may respond to the gender of their teacher by changing beliefs and behavior. The second is that teachers may discriminate by gender, either positively or negatively engaging students in the classroom depending on whether a student is of the same gender as the teacher (Hoffmann and Oreopoulos, 2009; Jones and Wheatley, 1990).

### **6.1 The effect of a role model on girls' beliefs and effort**

In this subsection, we conduct a test of the model's hypotheses that teacher-student gender match should positively change beliefs and investment behavior for low performing girls matched with female math teachers. In Table 4, we present results from estimating Equation 1, using students' response to the question "Do you think that boys are better at learning math than girls?" as the dependent variable. This serves a proxy for gender-stereotypical beliefs. The variable is coded as a dummy, equal to one if the respondent agrees that boys are better at learning math than girls. Our specification follows that used in column 5 of Table 3 (using grade-by-school fixed effects and the full battery of controls for students and teachers). We also control for students' math test scores, allowing us to compare, within those students with the same test score, how having a female teacher affects their beliefs about girls' ability to learn math.

The first column of Panel A of Table 4 shows that having a female math teacher significantly changes girls' gender-stereotypical beliefs. Being taught by a female math teacher reduces girls' probability of claiming to be inferior to boys in learning math by 6.8 percentage points, which is equivalent to 13% of the proportion of girls with male teachers who answer yes to the question. The second, third, and fourth columns of Panel A report estimates from the same cuts of the data.

As with the test score benefits of teacher-student gender match, the change in beliefs about

girls' math ability is concentrated among low performing girls. Being taught by a female math teacher reduces the probability that low performing girl students believe girls to be inferior to boys in learning math by 17.2 percentage points. This difference is equivalent to more than 30 percent of the overall proportion (0.55) of these girls answering "yes" to the question. The results in Column 3 and 4 suggest that the gender of the math teacher plays a smaller role in shaping normally or high-performing girls' gender-stereotypical beliefs. Importantly, low performing girls start from a lower average baseline belief: our estimates of  $\beta_1$  show that low performing girls as a group, no matter the gender of the teacher, are more likely than advantaged ones to believe that boys are better at learning math. In Appendix Table A.6, we show the same analysis for the above/below median cut of the data. Here the results are again similar - the below median group of girls assigned to a female math teacher is also 10 percentage points less likely to report negative gender stereotypical beliefs about girls' math ability. The coefficient for the above median group is an order of magnitude smaller and insignificant.

If parents who believe more in girls' math ability sort their children into classes with female teachers, this omitted variable could bias our estimates of the effect of teacher-student gender match upwards. To check for this possibility, we run the same regression as before, only replacing students' gender-specific beliefs about math learning ability as the dependent variable with their parents' beliefs. The logic behind this test is that parents' gender-stereotypical beliefs are unlikely to be related to the gender of math teachers other than through this sorting. This claim is based on two observations: one, that parents' priors are likely to be less flexible than those of their young children; and two, there are far fewer interactions between teachers and parents than there are between teachers and children. We present these results in Panel B of Table 4. We estimate that the effects of having a female math teacher on parents' gender-stereotypical beliefs are small, in magnitude, small relative to their associated standard errors, and, among the low performing group, more than an order of magnitude smaller than the effect on girls' beliefs.

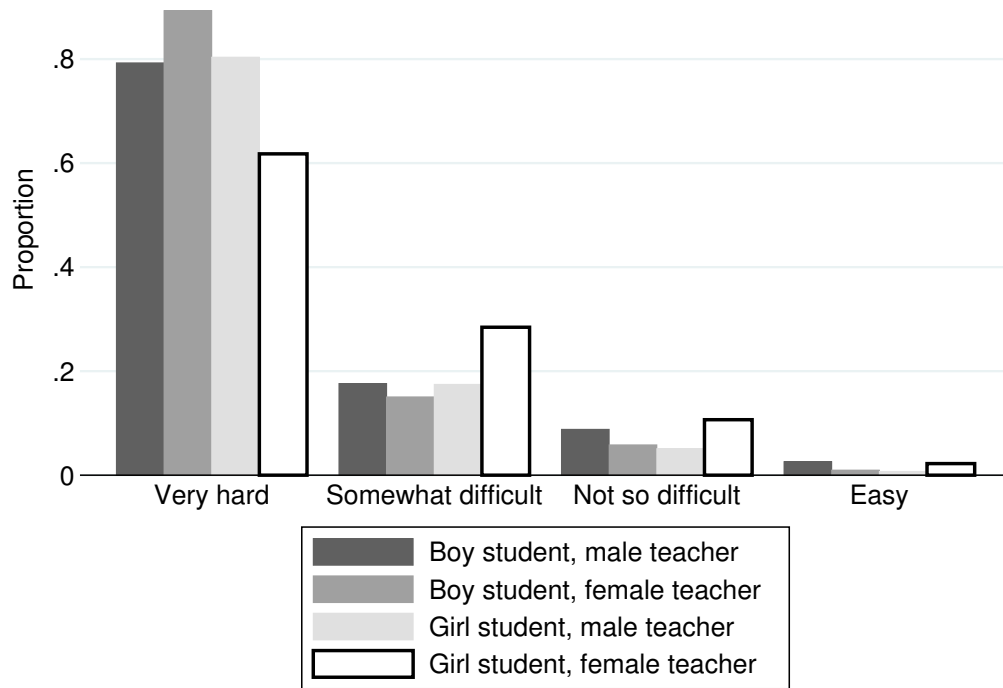
To corroborate this analysis, we look at the group of low performing girls' current beliefs about the difficulty of math. In Figure 2, we plot the proportion of these girls with each of four possible responses to the prompt "how difficult do you find your mathematics course at the moment." The potential responses are "very difficult", "somewhat difficult", "not so difficult", and "easy". These are plotted separately for low performing students in each of the four teacher-student gender pairings. Consistent with our previous results, those girls assigned to a female math teacher are nearly 20

Table 4: Effect of having a female teacher on student's self-concept

	All (1)	Low (2)	Average (3)	High (4)
<b>Panel A: Own beliefs</b>				
Female student ( $\beta_1$ )	-0.083*** (0.029)	0.239*** (0.045)	-0.032 (0.034)	-0.310*** (0.037)
Female teacher ( $\beta_2$ )	0.059* (0.031)	0.121 (0.114)	0.060 (0.039)	0.032 (0.041)
Female student * Female teacher ( $\beta_3$ )	-0.068** (0.035)	-0.172*** (0.069)	-0.031 (0.039)	-0.044 (0.046)
$\beta_2 + \beta_3$	-0.009	-0.051**	0.029	-0.013
Observations	8,151	826	4,403	2,888
<b>Panel B: Parent beliefs</b>				
Female student ( $\beta_1$ )	-0.050** (0.023)	0.117** (0.051)	-0.022 (0.026)	-0.200*** (0.040)
Female teacher ( $\beta_2$ )	0.042* (0.023)	-0.045 (0.109)	0.022 (0.034)	0.069* (0.037)
Female student * Female teacher ( $\beta_3$ )	-0.036 (0.027)	-0.008 (0.065)	0.014 (0.032)	-0.044 (0.048)
$\beta_2 + \beta_3$	0.007	-0.053	0.036	0.025
Observations	8,030	806	4,373	2,809

Notes: The dependent variable in Panel A is whether the student agrees with the statement “boys are better at learning mathematics than girls.” In Panel B, it is whether the parent agrees with the statement. The regression specification used is that laid out in Equation 1, adding a control for the student's math test scores. Point estimates and their precision are largely unchanged by removing this final control. Robust standard errors clustered at the school level are shown in parentheses. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

Figure 2: Low performing students' current perception of the difficulty of math, by gender of student and math teacher



Notes: This figure plots the response of low performing students to the prompt: how difficult do you find your mathematics course at the moment? This shows a clear rightward shift (towards lower perceived levels of difficulty in mathematics) for low performing girls assigned to a female teacher, relative to all other teacher-student gender pairings. The sample used to generate this figure is all low performing students in the estimation sample used for the regressions in Tables 3-4.

percentage points less likely to report that math is very difficult, and the graph shows an overall rightward shift, towards perceiving math to be less difficult, for low performing girls assigned to a female math teacher compared to all other groups. A similar picture appears in Appendix Figure A.1, which shows this analysis, only using the below median method of classifying low performers.

Another prediction of the model is that if aspirational distance decreases, investment should increase. We test this hypothesis using students' reported enrollment in math tutoring, and report the results in Table 5. We observe what the model predicts. We see in Panel A that the teacher-student gender match is associated with a 10 percentage point increase in proportion of low performing girls in math tutoring relative to boys. While the overall effect of a female teacher on girls' enrollment in tutoring ( $\beta_2 + \beta_3$ ) is negative, it is far more so for boys than for girls. It is possible that the motiva-

tional effect of teacher-student gender match improves the “quality” of girls’ studying, meaning that they learn more per given hour and do not require more hours. It may also be the case that the returns to investment in tutoring are not great in this context. In support of the latter claim, we note that  $\beta_2$  in column (1), the effect of a female teacher on low performing boys’ enrolling in tutoring in math, is negative and significant, while there is no attendant decrease in these boys’ performance (see the  $\beta_2$  coefficient for low performing students in Table 3). This suggests that low performing boys assigned to female teachers spend substantially less time in tutoring than those assigned to male teachers with no obvious negative consequences in terms of performance. This is consistent with either the time use data being unreliable or with a minimal impact of time spent in tutoring / doing homework on academic performance for low performing boys.

In Appendix Table A.5, we use data on weekly hours spent in tutoring, on homework related to tutoring, and homework assigned by the parent. These show a similar pattern - low performing girls spent more time in tutoring (Panel A), but no more time on homework (Panel B). In Panel A, teacher-student gender match is associated with a 3.3 hour boost in the hours low performing girls spend in after-school tutoring, with little movement for average or high performers. Similarly, low performing boys spend substantially less time on tutoring and homework. However, the results in Table A.5 are only suggestive, as the outcome variables are not specifically about math tutoring, but rather time spent in tutoring overall.

## 6.2 Teacher discrimination

Another possible explanation for the results in Section 5 is that teachers of a given gender may behave more or less favorably towards boys relative to girls (Beaman et al., 2009; Jones and Wheatley, 1990; Lavy, 2008; Lim and Meer, Forthcoming). In this subsection we test for evidence of such gender-specific discrimination.

The CEPS records students’ recall of whether their current math teacher asks them questions and praises them in the classroom, respectively. The data is collected using a four-point scale, ranging from 1 for “absolutely not” and 4 for “very often.” Table 6 shows results for estimating Equation 1 with these two measures of teacher behavior towards students as outcome variables. As in Tables 4 and 5, the first column shows estimates for the entire sample, and the following columns present subgroup-specific estimates for the low performing group and then the average

Table 5: Effect of having a female teacher on student's effort

	All (1)	Low (2)	Average (3)	High (4)
<b>Panel A: Attends tutoring in math</b>				
Female student ( $\beta_1$ )	0.015 (0.015)	-0.034 (0.031)	0.013 (0.018)	0.023 (0.028)
Female teacher ( $\beta_2$ )	-0.016 (0.023)	-0.175*** (0.064)	-0.018 (0.024)	-0.005 (0.040)
Female student * Female teacher ( $\beta_3$ )	0.038** (0.019)	0.094* (0.053)	0.037 (0.024)	0.012 (0.031)
$\beta_2 + \beta_3$	0.022	-0.082**	0.020	0.007
Observations	8,303	844	4,491	2,922
<b>Panel B: Attends math olympiad training</b>				
Female student ( $\beta_1$ )	-0.017** (0.007)	-0.016 (0.016)	-0.006 (0.010)	-0.020 (0.015)
Female teacher ( $\beta_2$ )	0.016 (0.015)	-0.074* (0.039)	0.019 (0.015)	0.036 (0.023)
Female student * Female teacher ( $\beta_3$ )	-0.007 (0.010)	-0.005 (0.025)	0.006 (0.013)	-0.027 (0.021)
$\beta_2 + \beta_3$	0.009	-0.079	0.025	0.009
Observations	8,303	844	4,491	2,922

Notes: The regression specification used here is given in Equation 1. The dependent variable in Panel A is whether a student attends any math tutoring. For Panel B, it is whether the student participated in advanced math / math olympiad tutoring. Robust standard errors clustered at the school level are shown in parentheses. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

and high performing students, respectively.

Panel A reports results on the effect of teacher-student gender match on the frequency of the student's interaction with the teacher. These results show that while female teachers are slightly more likely to ask students questions than male teachers ( $\beta_2$ ), there is no indication that female teachers favor girls in this behavior ( $\beta_3$ ). Examined by subgroup, there is again no evidence that low performing girls, who reap the largest benefits from teacher-student gender match, benefit from more interaction if their math teacher is also female.

Panel B of Table 6 shows estimates of the impact of teacher-student gender match on the frequency of the math teacher praising the student. Here we observe a pattern similar to Panel A: no evidence of preferential treatment towards same-gender students overall or differential praise to low performing girls, who gain the most in test scores and changes in beliefs from this match.

Teachers may also display gender-specific discrimination behavior through grading, as midterm tests are likely to be graded by teachers themselves, rather than by external reviewers (Ehrenberg et al., 1995; Fryer et al., 2008; Lavy, 2008). An alternative explanation for our results that we cannot fully rule out is that female teachers are particularly kind to low performing female students, which generates the gain in test scores and in beliefs about girls' math ability. We argue this possibility is unlikely, given that we see no evidence of differential treatment of these girls. Furthermore, an important difference between math tests and language tests is that language tests contain many subjectively assessed essay-type questions, while math tests are more often composed of questions which are assessed using (relatively) objective criteria, leaving less room for teachers' manipulation of test scores.

Existing literature also provides little evidence of test score manipulation based on the teacher-student gender match. For example, both Dee et al. (2016) and Diamond and Persson (2016) find that student gender is not correlated with the extent of score manipulation. Even if teachers discriminate in favor of girls (Lavy, 2008) or against them (Lavy and Sand, 2015), this gender discrimination will be absorbed by the student gender dummy in the regressions as long as the discrimination is not based on the teacher-student gender match, which our results on teacher behavior towards students based on teacher gender suggest to be the case. As a whole, we find scant evidence that math teachers appear to show gender-specific favoritism in their interactions with students in the classroom, suggesting the proposed discrimination channel, wherein math teachers favor same-gender students, is unlikely to drive the patterns we see in our main results.



Table 6: Teacher discrimination

	All (1)	Low (2)	Average (3)	High (4)
<b>Panel A: Frequency of being called on in class</b>				
Female student ( $\beta_1$ )	-0.108*** (0.035)	0.017 (0.127)	-0.054 (0.049)	-0.206*** (0.055)
Female teacher ( $\beta_2$ )	0.120* (0.065)	0.007 (0.241)	0.109 (0.067)	0.105 (0.094)
Female student * Female teacher ( $\beta_3$ )	0.020 (0.039)	-0.037 (0.183)	-0.003 (0.058)	0.105 (0.068)
$\beta_2 + \beta_3$	0.140* (0.065)	-0.030 (0.183)	0.105 (0.058)	0.210** (0.068)
Observations	8,268	839	4,481	2,917
<b>Panel B: Frequency of praise received</b>				
Female student ( $\beta_1$ )	-0.124*** (0.034)	0.015 (0.092)	-0.049 (0.054)	-0.185*** (0.057)
Female teacher ( $\beta_2$ )	0.096 (0.071)	0.128 (0.185)	0.110 (0.068)	0.073 (0.089)
Female student * Female teacher ( $\beta_3$ )	0.016 (0.040)	-0.147 (0.164)	0.002 (0.061)	0.083 (0.072)
$\beta_2 + \beta_3$	0.112 (0.040)	-0.019 (0.164)	0.112 (0.061)	0.156 (0.072)
Observations	8,277	840	4,492	2,914

Notes: The regression specification used here is given in Equation 1. For Panel A, the dependent variable is the response, on a four point scale from one being do not agree to four being fully agree, to the prompt “the teacher calls on me frequently.” Panel B’s dependent variable, on the same scale, is the response to the prompt “the teacher often praises me.” Robust standard errors clustered at the school level are shown in parentheses. All regressions control for the student’s math test scores, but the point estimates and their precision are largely unchanged by removing this control. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

We interpret these results as evidence in favor of one interpretation of the estimates presented in Section 5, specifically, that the large benefits that low performing girls in our sample reap from having a female math teacher (rather than a male one) are likely to come primarily from these girls' beliefs and behavior changing in response to the positive role model provided by a same-gendered teacher, and are unlikely to stem from female teachers discriminating against boys or in favor of girls.

## **7 Conclusion**

In this paper, we find that being assigned a female math teacher confers a series of positive effects on low performing female students. Teacher-student gender match is associated with a gain of 0.45 standard deviations for low performing girls on a math test, apparent in the empirical distribution of scores broken down by teacher-student gender pairings as well as in our regression results. We investigated several potential channels for this impact, and our findings suggest that the benefits stem primarily from a clear causal pathway predicted by psychological evidence and economic theory – having a female math teacher appears to provide a plausible, positive role model for low performing girls, and being assigned a female math teacher changes gender-stereotypical beliefs about math learning ability and the perceived difficulty of math. It is also associated with an increase in investment in human capital among these girls. We find no evidence for the competing explanation that female teachers favor girls over boys in the classroom, either by asking girls more questions or praising girls more frequently, particularly among those low performing female students who gain the most from teacher-student gender match. We demonstrate that there are no observable differences in teacher quality between female and male math teachers within the same grade within a school, suggesting that the gender-specific benefits we observe are unlikely to be associated with the differences in teacher quality between female and male teachers. Our evidence of the causal pathway from beliefs and aspirations to human capital investment and achievement adds to a growing body of research (Bernard et al., 2014; Bian et al., 2017; Genicot and Ray, Forthcoming) suggesting that the presence of a plausible example of success may be a key lever for changing beliefs, increasing effort in school, and, more broadly, an important input into the production of human capital, particularly among girls and other groups who may lack for a credible positive example of the returns to investment.

## REFERENCES

- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik**, “The effect of teacher gender on student achievement in primary school,” *Journal of Labor Economics*, 2015, 33 (1), 63–89.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, Petia Topalova et al.**, “Powerful Women: Does Exposure Reduce Bias?,” *Quarterly Journal of Economics*, 2009, 124 (4), 1497–1540.
- Beilock, Sian L, Elizabeth A Gunderson, Gerardo Ramirez, and Susan C Levine**, “Female teachers’ math anxiety affects girls’ math achievement,” *Proceedings of the National Academy of Sciences*, 2010, 107 (5), 1860–1863.
- Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Taffesse**, “The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia,” Technical Report, CEPR Discussion Papers 2014.
- Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian**, “Gender stereotypes about intellectual ability emerge early and influence children’s interests,” *Science*, 2017, 355 (6323), 389–391.
- Carman, Katherine Grace and Lei Zhang**, “Classroom peer effects and academic achievement: Evidence from a Chinese middle school,” *China Economic Review*, 2012, 23 (2), 223–237.
- Carrell, Scott E, Marianne E Page, and James E West**, “Sex and science: How professor gender perpetuates the gender gap,” *The Quarterly Journal of Economics*, 2010, 125 (3), 1101–1144.
- Cheryan, Sapna**, “Understanding the paradox in math-related fields: Why do some gender gaps remain while others do not?,” *Sex Roles*, 2012, 66 (3-4), 184–190.
- Cho, Insook**, “The effect of teacher–student gender matching: Evidence from OECD countries,” *Economics of Education Review*, 2012, 31 (3), 54–67.
- Dee, Thomas S**, “Teachers, race, and student achievement in a randomized experiment,” *Review of Economics and Statistics*, 2004, 86 (1), 195–210.
- , “Teachers and the gender gaps in student achievement,” *Journal of Human Resources*, 2007, 42 (3), 528–554.

- , **Will Dobbie, Brian A Jacob, and Jonah Rockoff**, “The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations,” *NBER Working Paper*, 2016.
- Diamond, Rebecca and Petra Persson**, “The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests,” *NBER Working Paper*, 2016, (w17727).
- Ehrenberg, Ronald G, Daniel D Goldhaber, and Dominic J Brewer**, “Do teachers’ race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988,” *Industrial & Labor Relations Review*, 1995, 48 (3), 547–561.
- Fryer, Roland G, Glenn C Loury, and Tolga Yuret**, “An economic analysis of color-blind affirmative action,” *Journal of Law, Economics, and Organization*, 2008, 24 (2), 319–355.
- Genicot, Garance and Debraj Ray**, “Aspirations and Inequality,” *Econometrica*, Forthcoming.
- Gunderson, Elizabeth A, Gerardo Ramirez, Susan C Levine, and Sian L Beilock**, “The role of parents and teachers in the development of gender-related math attitudes,” *Sex Roles*, 2012, 66 (3-4), 153–166.
- Hoffmann, Florian and Philip Oreopoulos**, “A professor like me the influence of instructor gender on college achievement,” *Journal of Human Resources*, 2009, 44 (2), 479–494.
- Holmlund, Helena and Krister Sund**, “Is the gender gap in school performance affected by the sex of the teacher?,” *Labour Economics*, 2008, 15 (1), 37–53.
- Jones, M Gail and Jack Wheatley**, “Gender differences in teacher-student interactions in science classrooms,” *Journal of Research in Science Teaching*, 1990, 27 (9), 861–874.
- Lai, Fang, Elisabeth Sadoulet, and Alain De Janvry**, “The contributions of school quality and teacher qualifications to student performance evidence from a natural experiment in Beijing middle schools,” *Journal of Human Resources*, 2011, 46 (1), 123–153.
- Lavy, Victor**, “Do gender stereotypes reduce girls’ or boys’ human capital outcomes? Evidence from a natural experiment,” *Journal of Public Economics*, 2008, 92 (10), 2083–2105.
- **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.

- , **Avraham Ebenstein, and Sefi Roth**, “The long run human capital and economic consequences of high-stakes examinations,” *National Bureau of Economic Research Working Paper 20647*, 2014.
- Lim, Jaegeum and Jonathan Meer**, “The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea,” *Journal of Human Resources*, Forthcoming.
- Muralidharan, Karthik and Ketki Sheth**, “Bridging education gender gaps in developing countries: The role of female teachers,” *Journal of Human Resources*, 2016, *51* (2), 269–297.
- Niederle, Muriel and Lise Vesterlund**, “Explaining the gender gap in math test scores: The role of competition,” *Journal of Economic Perspectives*, 2010, *24* (2), 129–144.
- Nixon, Lucia A and Michael D Robinson**, “The educational attainment of young women: Role model effects of female high school faculty,” *Demography*, 1999, *36* (2), 185–194.
- OECD**, *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, OECD Publishing, 2015.
- Paredes, Valentina**, “A teacher like me or a student like me? Role model versus teacher bias effect,” *Economics of Education Review*, 2014, *39*, 38–49.
- Rothstein, Jesse**, “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement,” *Quarterly Journal of Economics*, 2010, *125* (1), 175–214.
- Shih, Margaret, Todd L Pittinsky, and Nalini Ambady**, “Stereotype susceptibility: Identity salience and shifts in quantitative performance,” *Psychological Science*, 1999, *10* (1), 80–83.
- Spencer, Steven J, Claude M Steele, and Diane M Quinn**, “Stereotype threat and women’s math performance,” *Journal of Experimental Social Psychology*, 1999, *35* (1), 4–28.
- Wilson, W.J.**, *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press, 2012.
- Zhang, Hongliang**, “Identification of treatment effects under imperfect matching with an application to Chinese elite schools,” *Journal of Public Economics*, 2016, *142*.

## Appendix

Table A.1: Summary statistics for students

	All (1)	Female (2)	Male (3)	Difference (4)
Female (%)	48.71	-	-	-
Age	13.22	13.16	13.27	-0.10***
Minority (%)	11.31	11.78	10.86	0.92
Agricultural hukou (%)	48.44	47.55	49.28	-1.72
Father's years of education	10.69	10.75	10.62	0.13*
Father's highest level of schooling (%)				
Primary or below	13.77	13.46	14.07	-0.61
Middle school	41.14	40.93	41.33	-0.4
High school/technical school	25.43	25.04	25.79	-0.75
College or above	19.66	20.57	18.81	1.76**
Mother's years of education	9.97	10.08	9.87	0.20**
Mother's highest level of schooling (%)				
Primary or below	22.1	20.34	23.76	-3.42***
Middle school	38.11	39.7	36.59	3.12***
High school/technical school	22.92	22.58	23.25	-0.66
College or above	16.87	17.37	16.4	0.97
Number of siblings	0.69	0.75	0.64	0.11***
Household income "low"	18.11	16.97	19.18	-2.21***
Math test score	70.25	70.94	69.59	1.35***
Number of observations	8,345	4,065	4,280	-

Source: Baseline wave of CEPS, restricted to sample with random assignment of students to classes. Column 4 shows the gender differences in student characteristics with t-test results.

\*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

Table A.2: Summary statistics for teachers

	All (1)	Female (2)	Male (3)	Difference (4)
Female (%)	61.35	-	-	-
Age	37.94	36.95	39.5	-2.55**
Education level (%)				
Associate college or below	12.56	7.87	20	-12.13**
Part-time four-year university	34.78	33.07	37.5	-4.43
Full-time four-year university	48.79	54.33	40	14.33**
Master's degree or higher	3.86	4.72	2.5	2.22
Attended a normal university (%)	94.2	92.13	97.5	-5.37
Years of teaching experience	16.8	15.72	18.53	-2.81**
Holds a senior professional rank (%)	23.67	24.41	22.5	1.91
Won teaching award (%)				
At the province or national level	14.01	14.96	12.5	2.46
At the city level	43.96	42.52	46.25	-3.73
Observations	207	127	80	-

Notes: This table compares observable teacher characteristics across teacher gender. Column 4 shows the gender differences in teacher characteristics with t-test results. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

Table A.3: Regressions of gender on characteristics of math teachers, by school-level assignment mechanism

	Random assignment (1)	Non-random assignment (2)
Age of math teacher	-0.742 (2.264)	-5.500 (3.943)
Math teacher has university degree	0.032 (0.131)	0.417* (0.235)
Math teacher graduated from a teachers college	-0.065 (0.050)	-0.000 (0.000)
Math teachers years of teaching experience	0.581 (2.762)	-7.833 (5.332)
Math teacher has senior teaching qualification	-0.065 (0.114)	-0.083 (0.198)
qualification level of math teacher: senior + level 1	-0.065 (0.125)	-0.333* (0.193)
Math teacher received provincial or national teaching award	0.065 (0.114)	0.083 (0.153)
Math teacher received city-level teaching award or higher	-0.032 (0.131)	0.000 (0.219)
Math teacher received county-level teaching award or higher	0.065 (0.114)	-0.083 (0.198)
Math teacher received school-level teaching award or higher	0.000 (0.089)	-0.167 (0.171)
Number of teaching honors math teacher received in past 3 years	-0.038 (0.270)	-0.364 (0.290)
Math teacher received teaching honor in past 3 years	0.000 (0.154)	-0.273* (0.147)
Observations	215	223

Notes: Each cell reports the coefficient  $\gamma_1$  from the following regression  $y_i = \gamma_0 + \gamma_1 * gender_i + \eta_g + \varepsilon_i$ , where  $y_i$  is the outcome variable described in the first column and  $\eta_g$  is a grade-by-school fixed effect. This comparison, therefore, compares the characteristics of female and male math teachers within a grade, within a school. These regressions are estimated separately in columns (1) and (2). Robust standard errors are reported in parentheses below the coefficient.



Table A.4: Effect of teacher student gender match on math test score, cutting on median

	Below median (1)	Above median (2)
Female student ( $\beta_1$ )	1.054*** (0.420)	0.475** (0.215)
Female teacher ( $\beta_2$ )	1.629*** (0.550)	0.624** (0.278)
Female student * Female teacher ( $\beta_3$ )	1.168** (0.510)	0.058 (0.278)
$\beta_2 + \beta_3$	2.797***	0.682**
Observations	4,088	4,257

Notes: this table performs the same analysis as reported in Table 3, only cutting the data instead on whether the math test score was above/below the median.

Table A.5: Time use data

	All (1)	Low (2)	Average (3)	High (4)
<b>Panel A: Hours per week spent in tutoring</b>				
Female student ( $\beta_1$ )	-0.230 (0.286)	-2.472** (1.086)	0.016 (0.423)	0.029 (0.609)
Female teacher ( $\beta_2$ )	-0.477 (0.395)	-4.160** (1.870)	-0.361 (0.489)	-0.298 (0.601)
Female student * Female teacher ( $\beta_3$ )	0.664* (0.371)	3.383** (1.654)	0.289 (0.532)	0.246 (0.663)
Observations	8,057	805	4,365	2,849
<b>Panel B: Hours per week spent on homework (both tutor-assigned and parent-assigned)</b>				
Female student ( $\beta_1$ )	0.534* (0.324)	-0.853 (1.339)	1.064** (0.473)	0.052 (0.662)
Female teacher ( $\beta_2$ )	0.092 (0.472)	-3.114 (2.559)	0.410 (0.644)	-0.023 (0.503)
Female student * Female teacher ( $\beta_3$ )	0.637 (0.494)	1.240 (2.221)	0.428 (0.579)	0.715 (0.838)
Observations	8,034	805	4,350	2,840

Notes: The dependent variable in these regressions is given in the title of the Panel. The regression specification is given in Equation 1. Robust standard errors clustered at the school level are shown in parentheses. \*\*\* significant at 1 percent level, \*\* significant at 5 percent level, \* significant at 10 percent level.

Table A.6: Effect of teacher student gender match on beliefs, cutting on median test score

	All (1)	Below median (2)	Above median (3)
<b>Panel A: Own beliefs</b>			
Female student ( $\beta_1$ )	-0.083*** (0.029)	0.057* (0.034)	-0.232*** (0.033)
Female teacher ( $\beta_2$ )	0.059* (0.031)	0.087*** (0.035)	0.029 (0.045)
Female student * Female teacher ( $\beta_3$ )	-0.068** (0.035)	-0.110*** (0.043)	-0.019 (0.040)
$\beta_2 + \beta_3$	-0.009	-0.023**	0.010
Observations	8,151	3,962	4,189
<b>Panel B: Parent beliefs</b>			
Female student ( $\beta_1$ )	-0.050** (0.023)	0.070*** (0.029)	-0.171*** (0.026)
Female teacher ( $\beta_2$ )	0.042* (0.023)	0.075*** (0.027)	0.020 (0.036)
Female student * Female teacher ( $\beta_3$ )	-0.036 (0.027)	-0.064* (0.035)	0.000 (0.035)
$\beta_2 + \beta_3$	0.007	0.011**	0.020
Observations	8,030	3,908	4,122

Notes: this table performs the same analysis as reported in Table 4, only cutting the data instead on whether the math test score was above/below the median.

Table A.7: Effect of teacher student gender match on teacher interaction with student, cutting on median test score

	All (1)	Below median (2)	Above median (3)
<b>Panel A: Frequency of being called on in class</b>			
Female student ( $\beta_1$ )	-0.108*** (0.035)	-0.082 (0.053)	-0.136*** (0.046)
Female teacher ( $\beta_2$ )	0.120* (0.065)	0.050 (0.073)	0.190*** (0.069)
Female student * Female teacher ( $\beta_3$ )	0.020 (0.039)	0.037 (0.062)	-0.006 (0.057)
$\beta_2 + \beta_3$	0.140*	0.088	0.184**
Observations	8,268	4,039	4,229
<b>Panel B: Frequency of praise received</b>			
Female student ( $\beta_1$ )	-0.124*** (0.034)	-0.100** (0.049)	-0.150*** (0.048)
Female teacher ( $\beta_2$ )	0.096 (0.071)	0.120 (0.088)	0.100 (0.067)
Female student * Female teacher ( $\beta_3$ )	0.016 (0.040)	0.003 (0.057)	0.032 (0.059)
$\beta_2 + \beta_3$	0.112	0.123	0.132
Observations	8,277	4,042	4,235

Notes: this table performs the same analysis as reported in Table 6, only cutting the data instead on whether the math test score was above/below the median.

Table A.8: Effect of teacher student gender match on enrollment in tutoring, cutting on median

	All (1)	Below median (2)	Above median (3)
<hr/> <b>Panel A: Attends tutoring in math</b> <hr/>			
Female student ( $\beta_1$ )	0.015 (0.015)	0.026 (0.019)	0.004 (0.020)
Female teacher ( $\beta_2$ )	-0.016 (0.023)	0.003 (0.035)	-0.041* (0.024)
Female student * Female teacher ( $\beta_3$ )	0.038** (0.019)	0.017 (0.026)	0.054** (0.026)
$\beta_2 + \beta_3$	0.022	0.021	0.013
Observations	8,303	4,060	4,243
<hr/> <b>Panel B: Attends math olympiad training</b> <hr/>			
Female student ( $\beta_1$ )	-0.017** (0.007)	-0.016** (0.007)	-0.011 (0.013)
Female teacher ( $\beta_2$ )	0.016 (0.015)	0.017 (0.013)	0.019 (0.025)
Female student * Female teacher ( $\beta_3$ )	-0.007 (0.010)	0.001 (0.012)	-0.020 (0.016)
$\beta_2 + \beta_3$	0.009	0.018	-0.001
Observations	8,303	4,060	4,243

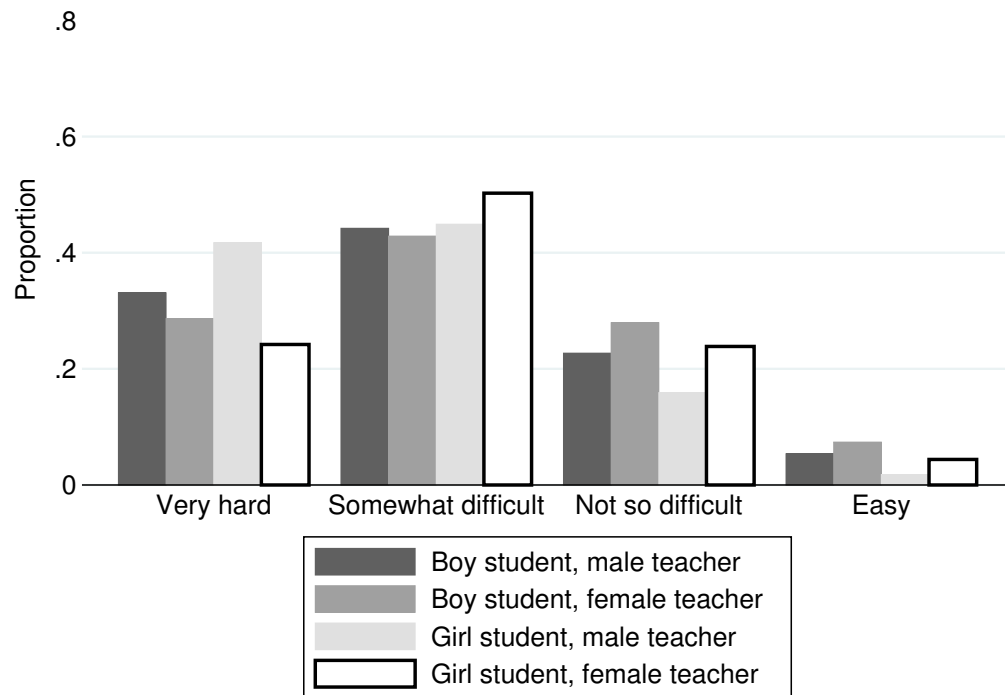
Notes: this table performs the same analysis as reported in Table 5, only cutting the data instead on whether the math test score was above/below the median.

Table A.9: Effect of teacher student gender match on student time use, cutting on median test score

	All (1)	Below median (2)	Above median (3)
<b>Panel A: Hours per week spent in tutoring</b>			
Female student ( $\beta_1$ )	-0.230 (0.286)	-0.200 (0.445)	-0.321 (0.430)
Female teacher ( $\beta_2$ )	-0.477 (0.395)	-0.235 (0.611)	-0.640 (0.466)
Female student * Female teacher ( $\beta_3$ )	0.664* (0.371)	0.967 (0.693)	0.373 (0.514)
$\beta_2 + \beta_3$	0.187	0.732	-0.267
Observations	8,057	3,900	4,157
<b>Panel B: Hours per week spent on non-class homework</b>			
Female student ( $\beta_1$ )	0.534* (0.324)	1.239** (0.533)	-0.150 (0.432)
Female teacher ( $\beta_2$ )	0.092 (0.472)	0.873 (0.786)	-0.710 (0.599)
Female student * Female teacher ( $\beta_3$ )	0.637 (0.494)	0.397 (0.782)	0.867 (0.600)
$\beta_2 + \beta_3$	0.729	1.269	0.158
Observations	8,034	3,895	4,139

Notes: this table performs the same analysis as reported in Table A.5, only cutting the data instead on whether the math test score was above/below the median.

Figure A.1: Effect of teacher-student gender match on student beliefs, for those below within-group median test score



Notes: this figure shows the same analysis as reported in Figure 2, only limiting the sample instead to those below the within-group median math test score.