COLUMBIA | SIPA
Center for Development Economics and Policy

CDEP-CGEG WORKING PAPER SERIES

CDEP-CGEG WP No. 43

Stereotypes, Role Models, and the Formation of Beliefs

Alex Eble and Feng Hu

April 2019

COLUMBIA | SIPA
Center on Global Economic Governance

Stereotypes, Role Models, and the Formation of Beliefs

Alex Eble and Feng Hu*

April 2019

Abstract

We study how information from stereotypes and role models influences children's beliefs, aspirations, and academic performance. We use a simple, stylized model of investment under uncertainty to formalize how female math teachers may affect the beliefs of students exposed to stereotypes about the math ability of each gender. It predicts differential effects by gender, and much larger effects among children who think they are of low ability in math. We exploit random assignment of students to classes in nationally-representative data from Chinese middle schools to test these predictions. We find that being assigned a female math teacher generates large gains in beliefs, aspirations, investment, and test scores for girls who perceive themselves to have low ability in math, generates moderate harms for boys with low perceived math ability, and has no gender-specific impact on these outcomes for non-low perceived ability children. We find no evidence that female math teachers teach differently than male teachers or give different praise or attention to low perceived ability students of different genders.

*Eble (corresponding author): Teachers College, Columbia University. Email: eble@tc.columbia.edu. Postal address: Department of Education Policy and Social Analysis. 525 W. 120th St. New York, NY 10027. Hu: School of Economics and Management, University of Science and Technology Beijing. Email: feng3hu@gmail.com. Author names are shown alphabetically; both authors contributed equally to this research. The authors are grateful to Anjali Adukia, Joe Cummins, John Friedman, Morgan Hardy, Asim Khwaja, Ilyana Kuziemko, Bentley MacCleod, Derek Neal, Randy Reback, Jonah Rockoff, Judy Scott-Clayton, and Felipe Valencia for generous input, as well as seminar audiences at Chicago, Columbia, Fordham, Stanford, the 2017 IZA transatlantic meeting, NEUDC, and PacDev. Funding: This work was supported by the National Natural Science Foundation of China (grant nos. 71373002, 71420107023). Declarations of interest: none. Keywords: gender; belief formation; stereotypes; human capital; cognitive skills; behavioral economics. JEL codes: D83; I24; J16.

1 Introduction

As a person goes through life, her beliefs are shaped by what she comes in contact with. These beliefs inform both her decisions on how to spend her resources, such as time and effort, and her aspirations for the future. This causal chain is particularly important in the formation of human capital (e.g., Becker 1975; Koch et al. 2015; Lybbert and Wydick 2016b). Furthermore, because the productivity of investment in human capital is partly determined by the amount of prior investment (Cunha and Heckman, 2007), this link from information to beliefs and on to decisions about time and effort in the early stages of life can have long-lasting consequences.

For example, if we erroneously tell a child she is of low ability in a given subject, she may reduce her relevant investment in that subject. When she takes her next test in the subject, she will get a lower score, reinforcing the erroneous message and potentially causing further reductions in effort and time. Exposure to stereotypes about ability by gender and ethnicity sends precisely this kind of message (Steele and Aronson, 1995; Steele, 2003; Bordalo et al., 2019). Recent evidence suggests that exposure to gender stereotypes affects the interests and time use decisions of both girls and boys as early as age seven (Bian et al., 2017) and that this may lead to underrepresentation of women and minorities in several scientific fields where such information persists (Leslie et al., 2015).

To counter these effects, many studies have found that matching students who face such stereotypes with teachers of the same identity - e.g., assigning girl students to female teachers and minority students to minority teachers - improves students' grades and persistence in school, and may change their expressed interests and choice of professional field (c.f., Dee 2004; Bettinger and Long 2005; Kofoed et al. 2017; Lim and Meer Forthcoming). Potential pathways for the benefits of teacher-student identity match include differential teaching methods by teacher identity, increased attention from teachers of the same identity, and the teacher serving as a role model for children facing stereotypes.

In this paper, we aim to advance understanding of how teacher-student identity match contributes to belief formation and shapes academic performance among children who are exposed to stereotypes. We study how children' beliefs, actions, and academic performance are affected by two interacting sources of information - stereotypes about gender-specific math ability and same-gendered math teachers. First, we use a simple, two-period model of consumption with uncertainty regarding returns to investment to predict who is most affected by teacher-student identity match in a subject with identity-specific stereotypes. We then take the model's predictions to rich data from Chinese middle schools which allow us to test these predictions and alternative hypotheses.

In our model, children are uncertain about their own ability in math and the returns to exerting math-related effort in school, and update their beliefs in response to the information they encounter. Teachers try to send the message that the student can achieve in math, and the signal they send is more credible if they share a gender with the child. The model's core prediction is that how much children update their beliefs in response to this message varies by two factors: one, the credibility of the signal the teacher sends, and two, the distance of this message from the students' prior about her own ability. This means that children who think they are of low ability are most likely to be affected by female math teachers.[1]

The model generates three specific, testable predictions. First, we predict that being assigned a female math teacher should generate gains for those girls who perceive themselves to be of low ability in math and whose beliefs about their own ability have been depressed by exposure to the stereotype that boys are better than girls at learning math. Second, we predict that being assigned a female math teacher should generate losses for low perceived ability boys, whose beliefs about their own ability have been inflated by exposure to the stereotype, and who may need a male math teacher to sustain their belief in their own ability to achieve in the subject. Third, we predict small or no effects for both boys and girls who are more certain they are of average or high ability.

We take these predictions to data from a nationally representative set of Chinese middle schools. This data has two features which facilitate our analysis. One, in this setting there is widespread belief among children that boys are better than girls in learning math, despite the fact that girls out-perform boys in mathematics. Two, there is random assignment of students to classrooms, which

[1] Bedard and Fischer (2019) find that students who perceive themselves to be low in the ability distribution are most affected by the threat of competition. See Jouini et al. (2018) for a more thorough theoretical treatment of these issues.

gives us random assignment of teacher gender to student gender. This source of plausibly exogenous variation has been used in several other studies of classroom configuration, both using data from Chinese middle schools (e.g., He et al. 2017; Gong et al. 2018) as well as in other countries (e.g., Lavy and Schlosser 2011; Lim and Meer 2017).

We estimate the effects of teacher-student gender match on girls' and boys' beliefs, behaviors, and academic outcomes, investigating heterogeneity in effect estimates by whether the student perceives her/himself to be of low ability in math prior to entering middle school. Our results match with the model's three predictions. First, girls who believe they are of low ability and who are assigned to female math teachers gain relative to all other low perceived ability children. They are 20 percentage points less likely to perceive their current math class as "very difficult" (from a baseline of 80%), are 11 percentage points less likely to aspire to jobs in the visual or language arts (baseline 23%), are nine percentage points more likely to enroll in mathematics tutoring (baseline 15%), and score 0.45 standard deviations (SD) better on a standardized math exam. Second, the beliefs and performance of low perceived ability boys assigned to female math teachers appear to deteriorate relative to those of boys assigned to male math teachers: they are 10 percentage points *more* likely to perceive math as very difficult, are eight percentage points *less* likely to enroll in math tutoring, and experience a non-significant (but relatively large) 0.15 SD drop in their math exam score. Third, we find no evidence of teacher-student gender match effects on the beliefs, aspirations, or test scores of girls or boys who enter middle school believing themselves to be of average or high ability in math.

Our results are robust across three different methods for defining low perceived ability, and persist when a subset of the students are tested again a year later. We conduct a battery of tests for the possibility that variation in teaching methods, aptitude, or effort between male and female teachers drive the effects we observe instead of the role model channel. We find no evidence that any of these differentially affect low perceived ability students. We use data on how frequently the teacher calls on and, separately, praises each student, to test for the possibility that female math teachers give extra attention to low perceived ability girls. Again, we find no evidence of such differential treatment.

Our work adds to two active literatures. The first is the budding set of studies on the formation of aspirations and beliefs and their role in forward-looking decisions, especially those related to human capital formation (e.g., Bernard et al., 2014; Macours and Vakis, 2014; Lybbert and Wydick, 2016a; Kofoed et al., 2017; Ross, 2017). We further this work by generating evidence that teacher-student identity match can help shape beliefs about ability and aspirations, particularly among children who enter a new environment (e.g., a new level of schooling) with a high level of uncertainty about their own ability and whose beliefs about themselves have been affected by exposure to stereotypes (Bian et al., 2017; Bordalo et al., 2019).

The second is the teacher-student identity match literature (e.g., Dee, 2004; Bettinger and Long, 2005; Carrell et al., 2010; Lim and Meer, 2017). The closest paper to ours is Gong et al. (2018), who use the same data we use to study the salutary effects of teacher-student gender match on *all* female students' academic outcomes and non-cognitive skills, aggregated across subjects (math, English, and Chinese). We advance on this previous work in two ways: one, we focus on the interplay between stereotypes, role models, and certainty of beliefs about ability, and two, we show, both conceptually and empirically, that the power of the messages that shared-identity teachers send to students about their potential to succeed in a subject may depend crucially on the distance of the message from the students' prior. This finding has important potential implications for how we assign students to teachers; it also complements recent work showing that shared-identity teachers serve as role models in other contexts, shaping career choices of stereotyped-against college students in the US (Carrell et al., 2010; Kofoed et al., 2017; Porter and Serra, 2017).

The rest of the paper proceeds as follows: Section 2 presents our conceptual framework, Section 3 briefly describes the setting we study, Section 4 describes our data sources and empirical strategy, Section 5 presents our main empirical results, Section 6 tests alternative explanations for our results and discusses the limitations of our study, and Section 7 concludes.

2 Conceptual framework

In this section, we use a simple two period model of consumption under uncertainty about the returns to investment to formalize our predictions of how different sources of information - stereotypes and teacher-student identity match - affect child beliefs, investment decisions, and performance.[2] We start with the premise that exposure to stereotypes can distort beliefs about oneself (e.g., Bian et al., 2017; Bordalo et al., 2019). Both across countries and in our Chinese data, girls express a disproportionate lack of confidence in their own ability in math as well as in the math ability of their gender; at least some of this difference is attributed to children's exposure to the stereotype that boys are better than girls at learning math (Beilock et al., 2010; OECD, 2015; Rodríguez-Planas and Nollenberger, 2018). The empirical literature in psychology demonstrates that these beliefs directly contribute to worse performance among women via two channels. First, anxiety because of "stereotype threat" (Shih et al., 1999; Spencer et al., 1999; Niederle and Vesterlund, 2010; Cheryan, 2012) could lead to lower performance on high-stakes math assessments, which would in turn affect later life outcomes. Second, negative gender norms may exert downward pressure on a child's beliefs about her returns to investment, causing girls to invest less effort, enthusiasm, and time in studying for math (Bian et al., 2017). As described in the introduction, this would lead to lower performance, confirming the once-erroneous content of the stereotype and, potentially, causing a subsequent further reduction in effort.

Encountering a female math teacher may affect children's views about their ability and the potential positive returns to their effort in math, particularly for children who were uncertain whether they *could* succeed in the subject. For girls, by virtue of shared gender, the female teacher provides a credible example of the returns to such effort (Carrell et al., 2010; Wilson, 2012; Genicot and Ray, 2017) which, in turn, may change girls' willingness to exert effort in the subject area (Nixon and Robinson, 1999; Beaman et al., 2009; Gunderson et al., 2012). In other words, the female teacher serves as a role model for girls. For boys, encountering a female math teacher in middle

[2] A version of this model was initially conceived during Eble's PhD research. Its genesis predates the empirical work here by several years and informed our plan of analysis.

school math - a stage when the subject matter gets much harder - could induce them to revise their beliefs about their ability in math downward, i.e., providing evidence that they are not necessarily good at math just because of their gender.

To formalize this intuition, we place our analysis in the context of a canonical two period model of consumption and savings.[3] Individuals face the following consumption problem:

$$\max_{s^i} E[u(y_1 - s^i, y_2 + r^i * s^i)] \tag{1}$$

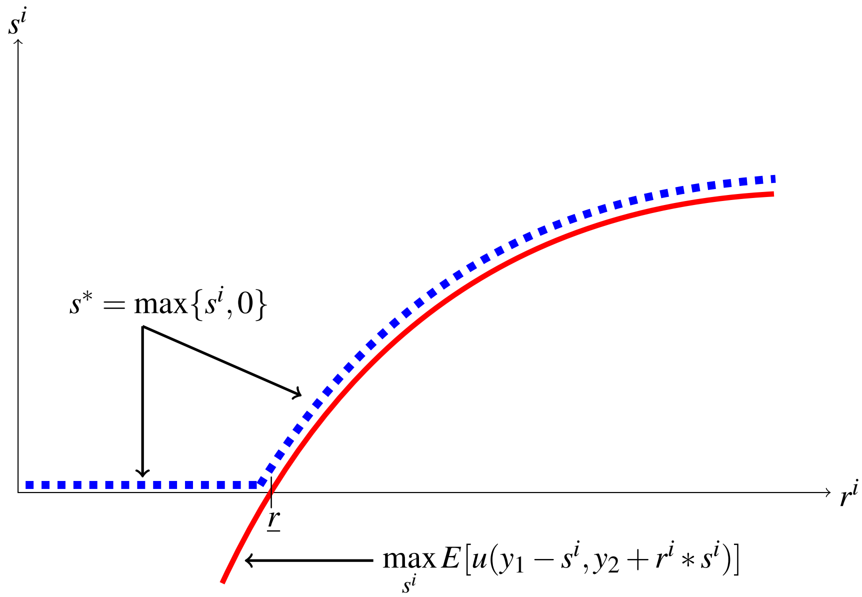
We index the individual in the superscript and time in the subscript. We assume utility from consumption is concave in both periods. We define y_t as income in period t, s^i as the savings of individual i in period 1, and r^i as individual i's belief about her return on saving in period 1, earned in period 2. Instead of modeling savings as money, here the savings technology is investment in human capital, a combination of effort and time exerted beyond the bare minimum in class, on homework, and in seeking out extra assistance via tutoring. We normalize this to zero when the student exerts the minimum possible effort. We assume r^i is a function of the individual's ability endowment and the informational environment the individual faces, which can include information gleaned from parents, peers, the media, and so forth. In Figure 1, we depict this static part of our model graphically, with s^* indicating the solution to the individual's optimization problem.

We introduce stereotypes through r^i. We assume there is part of the positive support of r, $[0, \underline{r})$, over which the interest rate does not justify investment. For some of the individuals whose perceived return on investment falls in this range, their low r^i is caused by incorrect information, that is, exposure to either stochastic shocks or group-based messages about ability such as stereotypes. These individuals will rationally but sub-optimally choose not to invest. Similarly, there will be some individuals who believe, incorrectly, that $r^i > \underline{r}$. Some of these individuals will hold this belief as a result of being told they are certain to have high returns simply because of their membership in a group (such as the male gender) alleged to be of superior ability.

We focus on how r^i changes in response to new information. How much r^i changes in response

[3] Jouini et al. (2018) formulate a richer model of many of these same issues.

Figure 1: Visual depiction of the model



to such information will depend on two parameters of the signal: one, its credibility, and two, the difference between the individual's prior and the new information provided by the signal. For girl students who perceive themselves to be of sufficiently low ability in math that they fall into the the $[0, \underline{r})$ part of the support of r^i, being assigned a same-gendered math teacher provides a signal of the potential for returns to investment in human capital that is both credible and novel. The signal is credible because of the teacher's shared gender. It is novel because it is far from these children's priors, which have been formed both by gender stereotypes and the low signals about ability received up to that point.

The converse of this prediction also holds. For boy students who perceive themselves to be of sufficiently low ability that they are near but not below $\underline{r}$, seeing a female teacher teaching a challenging math class provides a far less credible message that they can succeed in math. Because their beliefs about r^i have been artificially inflated by exposure to the stereotype that they are inherently better in math, receiving a less credible signal of their high ability and being confronted with difficult content (as happens at the transition to middle school math) may cause some boys to revise their beliefs downward to $r^i < \underline{r}$. For both boys and girls who think they are higher in the ability distribution, the signal sent by the teacher is closer to their prior, and therefore not informationally salient enough to change their beliefs.

This framework generates three main predictions that we take to our data. Prediction one is that girls assigned to a female math teacher will update their prior on their ability to productively invest in human capital more than boys assigned to either a female or male math teacher; this will in turn cause them to increase their investment and improve their performance. This is the standard prediction from previous studies of role models and teacher-student identity match (e.g., Bettinger and Long 2005; Porter and Serra 2017). Predictions two and three comprise our advance on the existing literature. Prediction two is that low-perceived ability girls will make much larger updates to their prior than high perceived ability girls in response to encountering a female math teacher. Prediction three is that being assigned a female math teacher, as opposed to a male one, will reduce low perceived ability boys' belief in their ability to productively invest in themselves. We present a more formal setup of these predictions, and the logic used to reach them, in Appendix C.

3 Setting

In our empirical work, we study a nationally representative sample of students who entered Chinese middle schools in the early 2010's. Since 1986, middle school has been compulsory in China. Since 2006, a separate law has banned tracking of students to different classes based on demonstrated ability or academic performance. There are currently 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.

Primary school graduates are assigned to a neighborhood middle school according to local educational authorities' regulations, e.g., districting. In the first system, they are then randomly assigned to classes by lottery or another quasi-random method.[4] In the second system, students

[4]For instance, according to alphabetical order by surname, i.e., every n^{th} student assigned to the n^{th} class.

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, requiring that the average quality of students be comparable across classes and the class not be bifurcated (Carman and Zhang, 2012).[5] Several recent papers exploit this random assignment of students to classes and provide explanations of the two different assignment mechanisms (Hu, 2015; He et al., 2017; Gong et al., 2018). Because a child's school is determined by place of residence and families are only allowed to send their children to schools in the area where their household residence permit was issued, there is little scope for sorting into schools/school districts with(out) random assignment.

In our empirical analysis, we exploit these two methods of assigning students to classes as providing potentially quasi-random matching of student gender to teacher gender. As described in Hu (2015) and Gong et al. (2018), who use the same data as we do, this system is not implemented with perfect fidelity. Despite the banning of class tracking, as students progress through middle school some schools may assign students to classes based on their academic performance in order to better prepare top students for the entrance examination; this practice is more common in the eighth and ninth grades than in the seventh. In this analysis, as in Hu (2015) and Gong et al. (2018), we restrict our attention to students randomly assigned to classes in the seventh grade and to students in those schools where random assignment of students to classes is maintained throughout middle school.

This paper was written concurrently with (and independently of) Gong et al. (2018), who use the first wave of the same dataset we use and a similar identification strategy, but who ask a separate research question using a different subsample of the data. Our paper studies how teacher-student identity match and gender stereotypes about math ability interact to shape children's beliefs about their own ability and their performance in mathematics. We use the guidance of our model to study how the information delivered by these two sources affects children differentially depending on the child's perception of her or his own ability. Gong et al. instead study the salutary effects of teacher-student gender match on *all* female students' academic outcomes and non-cognitive skills,

[5] In Appendix D, we provide a description of this type of assignment rule, borrowing from He et al. (2017).

aggregated across subjects (math, English, and Chinese). Their paper studies all children and does not differentiate between low and high perceived ability groups or between subjects which do or do not have stereotypes suggesting girls are of lower ability than boys. Using the first wave of the CEPS data, their sample (all classes, not just mathematics) and removing the perceived ability interaction terms, we are able to reproduce their findings. Finally, we differ from Gong et al. in our ability to show effects over time. Gong et al. use only the first wave of the CEPS; we also include the second wave of the data, allowing us to present how the effects we measure change as students progress through school.

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 empirical strategy we use, and Section 4.3 states and tests our main identifying assumptions.

4.1 Data sources

The main data source we use in 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.[6] The CEPS is a nationally representative longitudinal survey that aims to track middle school students through their educational progress and later labor market activities. Its sample was selected using 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. In each selected school, four classes were randomly chosen, two from the seventh grade and two from the ninth. All students in the selected classes were then surveyed. The CEPS uses five different questionnaires, administered to students, parents, homeroom (banzhuren) teachers, main subject (math, Chinese, and English)

[6] In Section 5.4 we provide supplementary analyses using the second wave, which only contains data for a subset of children.

teachers, and school administrators, respectively. It is China's first nationally representative survey targeting middle school students, comparable to the Adolescent Health Longitudinal Studies (AddHealth) in the U.S. and the National Education Panel Survey (NEPS) in Europe.

The CEPS contains rich demographic data on students, their families, and their teachers, as well as detailed information on students' beliefs, aspirations, and time use. It also 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. They are (relatively) low stakes exams, graded collectively by the math teachers in the student's grade. Although their grading is not always blinded, Gong et al. (2018) argue that blinded grading is common in these particular tests. In footnote 7, we make a slightly weaker argument: that low stakes math exam scores are unlikely to be substantially biased by teacher gender, and it is even less likely that they will be differentially biased for low perceived ability girls assigned to female teachers.[7]

The survey also collects data on the assignment mechanism used to assign students to classrooms, collected both from school principals and homeroom teachers.[8] The options are 1) tracking, 2) assignment according to students' household registration location, 3) either literally random assignment ("sui ji", meaning 'by chance') or according to the average-equilibrating algorithm described above, or 4) through other methods. About 85% of middle schools in our data 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.

[7] First, midterm exams in mathematics offer less scope for manipulation than English or Chinese because they are graded on more objective criteria (e.g., was the number produced the correct answer?). Second, in Section 6 we present evidence (Table 6) that female teachers do not favor girls or low perceived ability girls either with more opportunities to respond to questions or with more praise in the classroom, suggesting that female math teachers may also not favor low performing girls in grading.

[8] This data is self-reported. We argue that reporting bias in the assignment mechanism data is unlikely because the data collection process stresses the anonymity of the data (all identifying information is removed from the datasets released to scholars) and the data is collected by academics and graduate students, not government officials. We also limit the analysis to grades where both school principals and homeroom teachers report use of random assignment. Homeroom teachers are less likely than principals to face potential negative consequences of the school using a non-random assignment mechanism, and this restriction serves as a further check on the principal's self-report.

In our analysis, we will treat assignment to class as random for seventh graders in those schools reporting use of either purely random assignment or 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 after initial quasi-random assignment in the seventh grade. If this assumption is valid, our approach allows us to causally estimate the effect of teacher gender on student outcomes. This restriction is the same as in Hu (2015) and Gong et al. (2018), who also show the validity of this approach for causal inference using these data.

For brevity, we describe our summary results here and present tables in the appendix. Table A.1 presents summary statistics for students by gender for those students randomly assigned to classrooms, and Table A.2 shows summary statistics for teachers in the classrooms studied in Table A.1. Among the children in our sample, the average age of girls is lower than that of 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 at the school level.

Thirty-nine percent of the math teachers in our data are male, alleviating the challenge faced in Antecol et al. (2015) where there was an insufficient number of male teachers to draw strong conclusions from some of the comparisons made. Female math teachers are on average younger and less experienced than their male counterparts. Overall, female teachers appear to be slightly more qualified than their male counterparts in terms of education and proportion having won a teaching award at the province or national level.[9]

The significant differences in characteristics between girls and boys and between female and 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 next subsection, we show evidence that our empirical approach reduces the risk of potential bias

[9] 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 (the smallest of the three geographical units) are the least prestigious.

stemming from such heterogeneity between teachers, between schools. Specifically, our main specification compares outcomes of children within a school, within a grade, between children in a classroom with a female teacher and those in a classroom with a male teacher. The observed differences attenuate dramatically and cease to be significant at this level of comparison.

4.1.1 Classifying students as "low perceived ability"

We use three different specifications for classifying students as low perceived ability or not; our results are robust to choice of specification. For all three classifications, we assume that, in expectation, the r^i of students we classify as low perceived ability is closer to $\underline{r}$ than that of students we classify as high perceived ability. Our main specification uses the question CEPS asks students how difficult they found learning math in the sixth (and final) grade of primary school to proxy for students' perception of their ability.[10] Specifically, we classify those students who found learning math in the sixth grade to be "very difficult" as low perceived ability. We classify those who report sixth grade math to be "somewhat difficult," "not so difficult," or "easy" not to be of low perceived ability. In Table A.4 we show characteristics of students, by gender, for both of the perceived ability groups thus defined. Gaps between boys and girls described earlier persist across groups though, consistent with stereotypes, a higher proportion of girls perceive themselves to be low perceived ability than do boys (11.7% vs. 8.9%).

Our two alternative specifications for defining low perceived ability are as follows: in the first specification, we classify respondents who report sixth grade math to be either "very difficult" or "somewhat difficult" as low perceived ability; in the second, we classify those respondents as low perceived ability who score below the median level in the distribution of test scores within their teacher-student gender pairing cell (e.g., separately for boys assigned to female teachers, girls to female teachers, and so on). We summarize results using these two alternative specifications in the body of the paper next to the relevant analyses and report them in tabular form in the appendix.

[10]This is not intended to proxy for a student's actual ability, but rather, to (noisily) measure how able she thinks herself to be.

4.2 Empirical strategy

In this subsection we explain our estimating equation, which aims to identify the effects of being assigned a female math teacher on female and on male students differentially by their perceived ability. We estimate the following reduced form regression equation:

$$\begin{aligned} Y_{icgj} = \beta_0 + \beta_1 FS_{icgj} + \beta_2 FT_{cgj} + \beta_3 (FS_{icgj} * FT_{cgj}) + \gamma_0 LPA_{icgj} + \gamma_1 (LPA_{icgj} * FS_{icgj}) + \\ \gamma_2 (LPA_{icgj} * FT_{cgj}) + \gamma_3 [LPA_{icgj} * (FS_{icgj} * FT_{cgj})] + \beta_4 SC_{icgj} + \beta_5 TC_{cgj} + \eta_{gj} + \varepsilon_{icgj} \end{aligned} \tag{2}$$

The variables are defined as follows: Y_{icgj} denotes the outcome of interest for student i in class c of grade g in school j. FS_{icgj} is an indicator equal to one if student i is female, and FT_{cgj} is also an indicator, equal to one if the teacher in class c in grade g of school j is female. LPA_{icgj} is an indicator equal to one if the student perceives herself to be of low ability. SC_{icgj} is a vector of predetermined characteristics at the student level, TC_{cgj} is a similar vector for teachers, η_{gj} is a set of grade-by-school fixed effects, and ε_{icgj} is a robust standard error, clustered at the school level to allow for heteroskedasticity and arbitrary serial correlation across students within a given school.[11]

Unless otherwise specified, the controlled-for student-level characteristics determined prior to assignment of 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 (low income or not). The teacher-level predetermined characteristics include teacher 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 the city, province, or national level, respectively.

Intuitively, our estimation 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, by virtue of random assignment, the match of FS_{icgj} to FT_{cgj} is orthogonal to predetermined characteristics

[11] All of our results continue to hold if we instead cluster at the (less conservative) classroom level.

which may influence beliefs, investment, or achievement. We test this assumption later in this section.

All of our estimated coefficients display children's performance relative to non-low perceived ability boys assigned to a male teacher (the omitted category). The coefficients β_1, β_2, and β_3 indicate how all children with a certain characteristic (e.g., β_1: girls; β_2: children assigned to a female teacher; β_3: girls assigned to a female teacher) compare to this group. The coefficients γ_1, γ_2, and γ_3 indicate how low perceived ability children with these same characteristics (girls, students assigned to a female teacher, and the interaction) fare relative to low perceived ability boys assigned to male teachers.

To emphasize how we advance on the existing teacher-student identity match literature, we present our main results - for beliefs and for academic performance - sequentially. First, we show results estimated using the standard teacher-student gender match specification, i.e., without the low perceived ability interaction terms, as is done in most prior work (e.g., Muralidharan and Sheth, 2016; Lim and Meer, 2017. Second, we present results from the fully specified model, which includes the low perceived ability variable and its interactions. For the sake of (relative) brevity, we show estimates from only the fully specified model for our other results.

Our model generates clear predictions for three parameters. The first is for β_3 in the standard model and β_3 and γ_3 in the fully specified model, which we interpret as quasi-experimental estimates of the effect of being assigned a female math teacher on (low-perceived ability) girls relative to the effect for (low perceived ability) boys. This captures the effect of teacher-student gender match on the "gender gap" (Dee, 2007; Muralidharan and Sundararaman, 2011; Lim and Meer, 2017). Prediction 1 is that these coefficients should be non-zero and point in the direction of reducing the gender gap, e.g., positive for test scores and negative for perceived difficulty of math. The second parameter prediction is for γ_3, the effect on the gender gap for low perceived ability girls. Prediction 2 of our model is that γ_3 should be substantially larger in magnitude than β_3 in the fully specified model. Prediction 3 of our model pertains to γ_2. This coefficient captures the effect on all low perceived ability students of being paired with a female teacher, using those assigned to a male teacher as the comparison. By virtue of the inclusion of γ_3, γ_2 is also the entire effect of being assigned a female math teacher on low perceived ability boys. The model predicts γ_2 and γ_3 to differ in sign.

Note that if either prediction 1 or 3 is satisfied, it addresses concerns that "reversion to the mean" could be driving our results. The concern is that perhaps our low perceived ability students merely had a bad draw in their sixth grade test scores and this caused them to revise their beliefs about their ability downwards. Mean reversion predicts they would be likely to have a normal draw in seventh grade (Chay et al., 2005). This would raise their perceived ability and, possibly, test scores relative to others'. Such mean reversion, however, would lead to the prediction that all low perceived ability students should have a secular gain in test scores. Neither a positive γ_3 nor a difference in sign between γ_2 and γ_3 can be explained by reversion to the mean.

There are several parameters of ancillary interest that are derived from different combinations of the coefficients we estimate in equation 2, and we will explicitly address a few of these in our discussion of the empirical results. First, $\gamma_2 + \gamma_3$ yields the total effect on low perceived ability girls of being assigned a female math teacher relative to low perceived ability girls assigned a male teacher (that is, it is the sum of the effect of being assigned a female teacher on low perceived ability students and the effect of being assigned a female teacher specific to low perceived ability girls). Second, $\beta_3 + \gamma_3$ yields the total effect of teacher-student gender match on the gender gap for low perceived ability girls (i.e., making the comparison group all boys, not only low perceived ability boys).

4.3 Identification

If our assumption of orthogonality is satisfied, estimating equation 2 using OLS should recover unbiased estimates of these parameters. To test this assumption - that within a grade within a given school, the match of student gender to teacher gender is as good as random - we follow Antecol et al. (2015), regressing math teacher gender on the same set of observable, predetermined student and family characteristics described above that we control for in our main empirical specification.[12] We conduct two regressions - one without any fixed effects, and a second with the grade-by-school fixed effects we use in our main empirical specification. For each regression we present coefficient estimates and report the F-statistic and p-value from a Wald Test of the joint significance of the regressors. We present these results in columns 1 and 2 of Table 1. With the inclusion of grade-by-school fixed effects, our F-test fails to reject the null that the regressors are together not significant predictors of teacher gender (column 2). Though one of the twelve individual coefficients is statistically significant, this is consistent with statistical chance. These results support our main identifying assumption that students' observable predetermined background characteristics are balanced along the gender of math teachers within the same grade in a given school.[13] While we cannot rule out the possibility that in some cases influential parents or individuals successfully lobbied to be placed with a certain teacher, we conclude from these results that such non-random matching of teachers to children is unlikely to be common enough to substantially bias our estimates.

Another descriptive comparison of interest is teacher quality across genders. This paper aims to investigate the effect of female math teachers on student achievement. To ensure that we are isolating the effect of gender, we need to establish whether male and female teachers differ on observable characteristics, such as teaching skill, which could drive any effects we measure (Cho, 2012; Antecol et al., 2015). To do so, we conduct an empirical test similar to that in Table 1, only with the analysis at the teacher level. The predetermined characteristics we include on the right hand side are age, a dummy for having earned a full-time bachelor's degree or higher qualification, a dummy for having attended a "normal" university (i.e., a university specializing in teacher training), years of teaching experience, and two dummies for having won a teaching award at two different levels, respectively. After conditioning on grade-by-school fixed effects, we again fail to reject the null that within a grade within a school, these characteristics are not jointly predictive of

[12] This method is also discussed in Hansen and Bowers (2008) and Bruhn and McKenzie (2009).

[13] Though we would like to conduct a synthetic randomization test, as in Carrell and West (2010) and Kofoed et al. (2017), we lack pre-assignment performance data. As a result, we cannot further test our assumption that class assignment is orthogonal to student aptitude.

the teacher's gender. Table A.3 reports these estimation results.

As we rely on teachers' and principals' reports of whether they use tracking or random assignment, it may also be the case that some schools which report using random assignment in fact use tracking. Deliberate misreporting of tracking as "random" would bias upward our estimates of the effect of female teachers on the best students (i.e., β_3) and bias downward the effect on worse students (γ_3), who are less likely to be assigned to "good" teachers under a tracking system in which the administrators are seeking to maximize the performance of the best students. Bias from this misreporting would push our coefficient estimates in the opposite direction of our framework's main predictions.

We note that our perceived ability data is observed at the same time as all of the other data, specifically, after teacher assignment. It is possible, therefore, that teacher gender may affect a child's report of the difficulty she had in primary school math, possibly in a way that is correlated with controlled-for predetermined characteristics such as gender. To test for this possibility, we run the same regressions of teacher gender on our list of predetermined characteristic controls, only restricting our analysis to low perceived ability students. We show our results in columns 3 and 4 of Table 1. The general pattern is the same as that for the entire sample - after controlling for grade-by-school fixed effects, only one of the 12 estimated coefficients is statistically significant and we fail to reject the null that these characteristics are jointly insignificant predictors of teacher gender. In other words, we find no evidence that predetermined student characteristics impact a child's likelihood of reporting low perceived ability (i.e., presence in the low perceived ability sample) in a way that is correlated with the gender of their math teacher.

It is also possible that the determinants of perceived ability differ between boys and girls in a way that may predict their test scores. If this were true, it would influence our interpretation of γ_3. To examine this possibility, we regress test scores on the vector of student-level predetermined characteristics and, using these coefficients, generate a predicted test score for each student. In Figure A.1, we plot these predicted test scores separately for boys and girls in each of the two perceived ability groups. These plots show no evidence of differences in the distribution of predicted

Table 1: Test for randomization

	Full sample		*Low perceived ability*	
	(1)	(2)	(3)	(4)
Number of siblings	-0.021	-0.006	-0.026	0.001
	(0.016)	(0.006)	(0.025)	(0.013)
Household is poor	-0.053	0.005	-0.100**	0.014
	(0.033)	(0.013)	(0.046)	(0.026)
Female	0.000	0.003	-0.078**	-0.015
	(0.012)	(0.005)	(0.037)	(0.016)
Age	-0.040	-0.011**	-0.071***	-0.006
	(0.025)	(0.005)	(0.028)	(0.007)
Ethnic minority	-0.150*	0.013	-0.109	0.026
	(0.089)	(0.018)	(0.099)	(0.023)
Holds agricultural hukou	-0.057*	-0.010	-0.112*	-0.042
	(0.032)	(0.013)	(0.057)	(0.032)
Mother's education level				
Middle school	0.125***	0.009	0.140***	-0.008
	(0.031)	(0.013)	(0.046)	(0.022)
High/technical school	0.112***	0.003	0.115	0.043
	(0.035)	(0.013)	(0.074)	(0.038)
College or above	0.139***	0.005	0.066	-0.069
	(0.041)	(0.015)	(0.102)	(0.065)
Father's education level				
Middle school	0.038*	-0.010	0.065	-0.012
	(0.022)	(0.009)	(0.040)	(0.027)
High/technical school	0.022	0.000	0.018	-0.041
	(0.030)	(0.014)	(0.063)	(0.045)
College or above	0.051	0.010	0.267***	0.149***
	(0.036)	(0.017)	(0.075)	(0.061)
Low perceived ability in math	-0.058*	-0.015		
	(0.033)	(0.018)		
Grade-by-school fixed effects		X		X
Number of observations	8,294	8,294	850	850
R-squared	0.08	0.66	0.18	0.85
Joint test F-statistic	3.21	0.97	14.27	1.55
[p-value]	[0.00]	[0.48]	[0.00]	[0.12]

Notes: This table shows results from four separate regressions of math teacher gender (=1 if female) on the set of independent variables listed in the first column, following the test in Antecol et al. (2015). Columns 1 and 2 show estimates generated using the entire sample and columns 3 and 4 show estimates generated using the low perceived ability group only.

test scores between genders in either group.

5 Main empirical results

In this section, we present results from applying our empirical strategy to the CEPS data. First, we estimate the impact of teacher-student gender match in mathematics on children's beliefs and aspirations. We then look at how this match affects investment in human capital and performance in mathematics.

5.1 Beliefs and aspirations

In this subsection, we conduct tests of the model's prediction that being assigned a female math teacher should positively affect beliefs and aspirations for low perceived ability girls and negatively affect them for low perceived ability boys. We investigate the impact of teacher gender on two belief variables: perceived difficulty of current math class and the careers to which students aspire. Our specification follows equation 2, 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 changes in beliefs while controlling for performance (our results are robust to removing these from the specification).

For the analysis of perceived difficulty, we use students' response to the prompt "how difficult do you find your *current* math class to be?"[14] The potential responses are "very difficult," "somewhat difficult," "not so difficult," and "not difficult at all", and we code the variable, as we do with perceived ability, with an indicator equal to one if the response is "very difficult." To study the impact of teacher-student gender match on aspirations, we use children's response to the prompt "what job would you most like to do when you grow up?" There are several possible responses to the question.[15] We investigated two potential outcomes: one, on the lower end of aspirational

[14]Recall that the baseline perceived ability question asked about the child's experience in the sixth grade; this question refers to the child's current experience in either the seventh or ninth grade.

[15]The options are 1. Government Official, 2. Business manager, 3. Scientist/engineer, 4. Teacher/doctor/lawyer, 5. Designer, 6. Artist/actor, 7. Athlete, 8. Skilled worker, 9. Other, 10. Don't care, 11. Don't know.

change, we created a variable for whether or not the child aspired to jobs traditionally associated with women[16]; two, we created a variable for whether or not the child aspired to jobs in the STEM fields.

We present our results in Table 2. First, we show results with the standard teacher-student gender match specification (as in, e.g., Muralidharan and Sheth, 2016), in columns 1 and 3. Second, we present results from the fully specified model, which includes the low perceived ability variable and its interactions, in columns 2 and 4. In column 1, being assigned a female math teacher is associated with an eight percentage point drop in the likelihood of girls perceiving math to be difficult. In column 2, adding in the low perceived ability interactions, we see that all of these benefits appear to accrue to low perceived ability girls assigned to female math teachers: being taught by a female math teacher reduces low perceived ability girls' probability of perceiving math as "very difficult" (γ_3) by 20 percentage points. While the estimated effect for non-low perceived ability girls assigned to a female teacher (β_3) is the same sign as for the low perceived ability girls, it is an order of magnitude smaller and not statistically significant. Being assigned a female math teacher is also associated with a 10 percentage point increase in low perceived ability boys' perceived difficulty of math (γ_2). These results accord with predictions 1-3 from the model.

We also present results visually in Figure 2. In this figure, we plot the distribution of perceived difficulty of the current math class for each possible teacher-student gender pairing. In Panel A, we restrict the sample to low perceived ability children. This shows the same pattern as the coefficients - low perceived ability girls assigned to a female teacher are at least 20 percentage points less likely to perceive math to be very difficult than any other group, and low perceived ability boys assigned to female math teachers are at least 10 percentage points more likely to find math very difficult than any other group. In Panel B, we show the same results for the non-low perceived ability group. Consistent with our framework's predictions, Panel B shows no detectable difference in the perceived difficulty of the current math class between non-low perceived ability girls assigned to

[16]In the raw data, women are most likely to choose jobs in the language and visual arts (designer; artist/actor), and we generate a variable "aspires to jobs in art and design" equal to one if the job aspired to is one of these and equal to zero otherwise.

female teachers and all other groups.

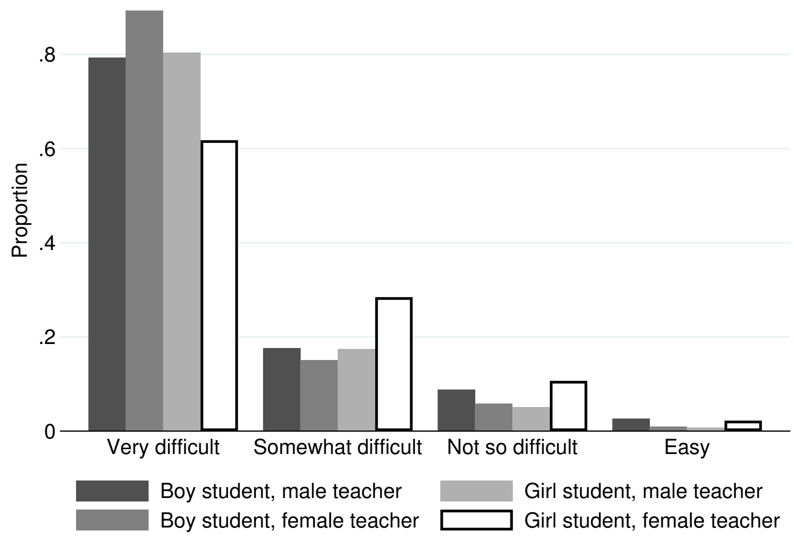
Table A.5 shows these results for the alternative specification of low perceived ability, where we observe an 11 percentage point decrease in the perceived difficulty of mathematics for low perceived ability girls thus defined. In column 1 of Table A.6, we present estimates generated using students below the within-group median test score instead of the low perceived ability group. We observe below-median girls assigned to a female teacher are 7.8 percentage points less likely to find math very difficult. While these are smaller than the coefficients generated using the original specification of low perceived ability, the estimates retain both their predicted sign and statistical significance. Graphically, Figure A.2 gives the below-median analogue to Figure 2 and displays a similar pattern.

Table 2: Effects on beliefs and aspirations

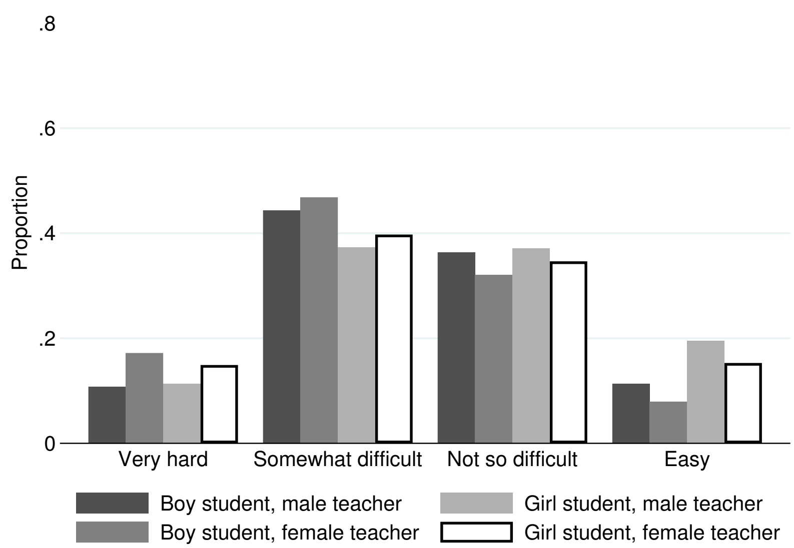
	Current math class perceived as very difficult (1)	(2)	*Aspires to jobs in art, art, design, or acting* (3)	(4)
Girl x female teacher x low perceived ability	—	-0.205***	—	-0.110**
		(0.057)		(0.056)
Female teacher x low perceived ability	—	0.100**	—	-0.031
		(0.046)		(0.034)
Girl x female teacher	-0.078***	-0.037	-0.011	0.008
	(0.019)	(0.017)	(0.018)	(0.019)
Girl x low perceived ability	—	0.046	—	0.105***
		(0.042)		(0.034)
Girl	0.092***	0.051***	0.202***	0.184***
	(0.016)	(0.013)	(0.016)	(0.017)
Female teacher	0.030	0.010	0.005	0.004
	(0.020)	(0.017)	(0.018)	(0.018)
Low perceived ability	—	0.451***	—	-0.003
		(0.037)		(0.026)
Mean for non-LPA boys	0.122		0.104	
Number of observations	8,276		8,213	

Notes: The regression specification used is given in equation 2, adding a control for the student's math test scores. Point estimates and their precision are largely unchanged by removing this final control. Both dependent variables are coded as (0 = No , 1 = Yes). Robust standard errors clustered at the school level are shown in parentheses. *p < 0.1, **p < 0.05, ***p < 0.01. Variation in the number of observations here and in subsequent tables stems from missing values in the dependent variable. Results are robust to restricting the sample to only observations with no missing dependent variables.

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



Panel A: Low perceived ability students



Panel B: Non-low perceived ability students

Notes: This figure plots the students' response to the prompt: "how difficult do you find your current mathematics course to be?" by the gender of the student and teacher. Panel A shows a clear rightward shift (towards lower perceived levels of difficulty in mathematics) for low perceived ability girls assigned to a female teacher relative to all other teacher-student gender pairings. Panel B shows no detectable difference in the perceived difficulty of the current math class among non-low perceived ability students between girls assigned to female teachers and all other pairings.

In columns 3 and 4 of Table 2, we present estimates of the effect of being assigned a female math teacher on students' aspirations to jobs traditionally associated with women. In column 3, the coefficients suggest that while girls are more likely to aspire to these jobs, there is no detectable effect of being assigned a female math teacher on girls' aspirations. In column 4, however, we estimate that for low perceived ability girls, being assigned a female math teacher is associated with an 11 percentage point decrease in aspiring to traditionally female jobs. The effects of being assigned a female teacher on all other groups (low perceived ability boys, all other boys and girls) are at least an order of magnitude smaller and insignificant. In Table A.5 the coefficient estimate using the alternative specification of low perceived ability has the predicted sign but is not statistically significant, and in Table A.6 we see no effect on aspirations for the below-median girls assigned to female teachers.

We find no effects on girls' aspirations to jobs in STEM fields, either for low perceived ability children or the group as a whole, and so do not present the results. One potential explanation for this is that because of the higher selectivity of the STEM fields, low ability children in our study are not on the margin of aspiring to work in those fields. The model also suggests that it may be much harder for teachers to change the beliefs of non-low perceived ability children who might be closer to this other margin.

5.2 Investment in human capital

We next conduct a series of tests of the model's prediction that teacher-student gender match should positively change investment behavior for low perceived ability children. We test this using four different dependent variables: students' reported enrollment in math tutoring, their total hours in tutoring (including, but not only, math tutoring), their hours spent on homework, and their enrollment in math olympiad tutoring. We give these results in Table 3.

Estimates presented in column 1 suggest that, for low perceived ability girls, teacher-student gender match is associated with a 9.1 percentage point increase in enrollment in math tutoring (significant at the 10% level). Low perceived ability boys assigned to female teachers, on the other hand, spend substantially less time in tutoring than those assigned to male teachers. These estimates of γ_2 and γ_3 also agree with our model's predictions, though it is worth noting that the total effect of being assigned a female teacher for low perceived ability girls, $\gamma_2 + \gamma_3$, is a 0.9 percentage point difference (essentially zero), meaning that they are on par with low perceived ability students assigned to male teachers. In Section 6, we explore the possible sources (i.e., parents, teachers, or children) of this change in investment. Also, and again as predicted, we see a much smaller and statistically insignificant estimate of β_3, the [girl x female teacher] coefficient.

In columns 2 and 3, we present estimates of the effect of teacher-student gender match on time use, first for weekly hours spent in tutoring, then for hours per week spent on homework. In column 2, we find that teacher-student gender match generates a statistically significant increase in the hours spent in tutoring for low perceived ability girls (three hours per week). We estimate that being paired with a female math teacher leads to a non-significant decrease in hours spent in tutoring for low perceived ability boys (1.5 hours). These results are only suggestive, however, as the time use data is not specifically about math tutoring, but rather time spent in tutoring overall. In column 3, we see no significant effect on hours spent on homework for either group.

In column 4, we estimate the effect of being assigned a female math teacher on enrollment in math olympiad tutoring. This tutoring is designed for students who aim to develop advanced math skills. Since the low perceived ability girls also have lower math test scores than their peers, it is unlikely that the differences in beliefs apparently induced by a female math teacher would lead to substantial gains in olympiad tutoring, which is targeted at students of relatively higher ability. On the other hand, if role models also affect beliefs and behavior at the higher end of the perceived ability spectrum, we may find an impact on olympiad tutoring for higher perceived ability girls. In line with what our model predicts, we see no significant effect of being assigned a female math teacher on enrollment in math olympiad tutoring among girls at any perceived ability level. Finally, consistent with the posited negative influence of anti-girl stereotypes in math, girls in our sample are 30% (1.8 percentage points) less likely to enroll in math olympiad tutoring than boys despite girls' superior performance on mathematics examinations.

Table 3: Effects on investment in human capital

	(1) Enrolled in math tutoring	(2) Hours in tutoring	(3) Hours spent on homework	(4) Math olympiad tutoring
Girl x female teacher x low perceived ability	0.091*	3.057***	0.392	0.000
	(0.052)	(1.253)	(1.595)	(0.024)
Female teacher x low perceived ability	-0.082**	-1.548	0.687	-0.014
	(0.036)	(0.996)	(1.353)	(0.022)
Girl x female teacher	0.027	0.262	0.516	-0.006
	(0.019)	(0.363)	(0.519)	(0.011)
Girl x low perceived ability	-0.054	-2.203***	-1.044	-0.001
	(0.035)	(0.933)	(1.252)	(0.018)
Girl	0.022	0.080	0.716*	-0.018**
	(0.016)	(0.295)	(0.403)	(0.008)
Female teacher	-0.012	-0.262	0.095	0.018
	(0.023)	(0.403)	(0.467)	(0.016)
Low perceived ability	0.041	1.231	0.081	0.008
	(0.026)	(0.832)	(1.052)	(0.016)
Mean for non-LPA boys	0.210	4.046	5.545	0.063
Number of observations	8,257	8,019	7,995	8,257

Notes: The regression specification used here is given in equation 2. The dependent variable is given in the column headings. Dependent variables in columns 1 and 4 are coded as (0 = No , 1 = Yes). Robust standard errors clustered at the school level are shown in parentheses. *p < 0.1, **p < 0.05, ***p < 0.01.

5.3 Academic performance

In this subsection we examine the effect of teacher-student gender match on students' performance in mathematics. We present our main results in Table 4. The first column shows the estimates generated from a version of equation 2 with no low perceived ability variables. The second column shows estimates generated using the main specification, including the low perceived ability variable and its various interaction terms on the right hand side, as specified in equation 2.

Our results bear out the predictions of our model. In column 1, the specification without the low perceived ability girl interaction terms, we see a positive but not statistically significant effect of teacher-student gender match on all girls: $\beta_3 = 0.093$ SD ($\sigma = 0.063$). This point estimate is well within the range of estimates generated in previous work (e.g., Dee, 2007; Muralidharan and Sundararaman, 2011; Lim and Meer, 2017). In column 2, when we add the low perceived ability interaction terms, we see again that the benefits estimated in column 1 appear to accrue entirely to low perceived ability girls. Being assigned a female math teacher increases the math test scores of low perceived ability girls by approximately 0.45 SD. In line with prediction 2 from our conceptual framework, girls who do not perceive themselves to be low ability appear to gain no gender-specific benefit from being assigned a female teacher (the coefficient estimate for β_3 is less than 0.01 SD). Consistent with prediction 3 and the patterns shown in the previous subsections, we also see some evidence that low perceived ability boys' test scores decline, though the estimate is not significant and is a third of the size of the estimated effect for low perceived ability girls. While our estimate of 0.45 SD is quite large, it is estimated for a subgroup that our conceptual framework predicts is particularly likely to benefit from teacher-student gender match. Other work evaluating interventions in developing countries that targeted either low performers or those in particularly needy regions finds similarly large effects (Banerjee and Duflo, 2007; Burde and Linden, 2013).

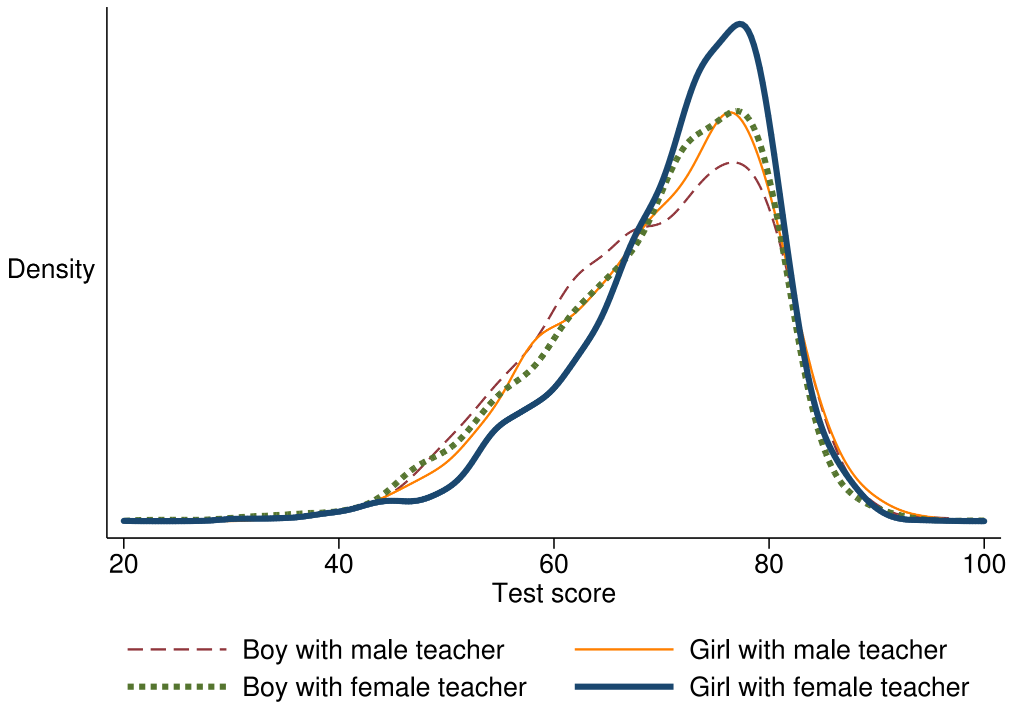
Next, in Figure 3, we show a kernel density plot of math test scores for the four different teacher-student gender pairings ($G^f : T^f$, $G^f : T^m$, $G^m : T^f$, and $G^m : T^m$). Girls assigned to a female math teacher outperform all other pairings, but only in (roughly) the left half of the distribution. A Kolmogorov-Smirnov test rejects the equality of the $G^f : T^f$ distribution from the

Table 4: Effects on math test scores

	(1)	(2)
Girl x female teacher x low perceived ability	—	0.446***
		(0.166)
Female teacher x low perceived ability	—	-0.147
		(0.129)
Girl x female teacher	0.093	0.007
	(0.063)	(0.054)
Girl x low perceived ability	—	-0.019
		(0.125)
Girl	0.068	0.125***
	(0.057)	(0.049)
Female teacher	0.155**	0.185***
	(0.074)	(0.068)
Low perceived ability	—	-0.806***
		(0.084)
Mean for non-LPA boys	7.024	7.024
Number of observations	8,345	8,294

Notes: The dependent variable is the student's math test score, shown here with the standard deviation standardized to 1 for comparability with other relevant studies. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

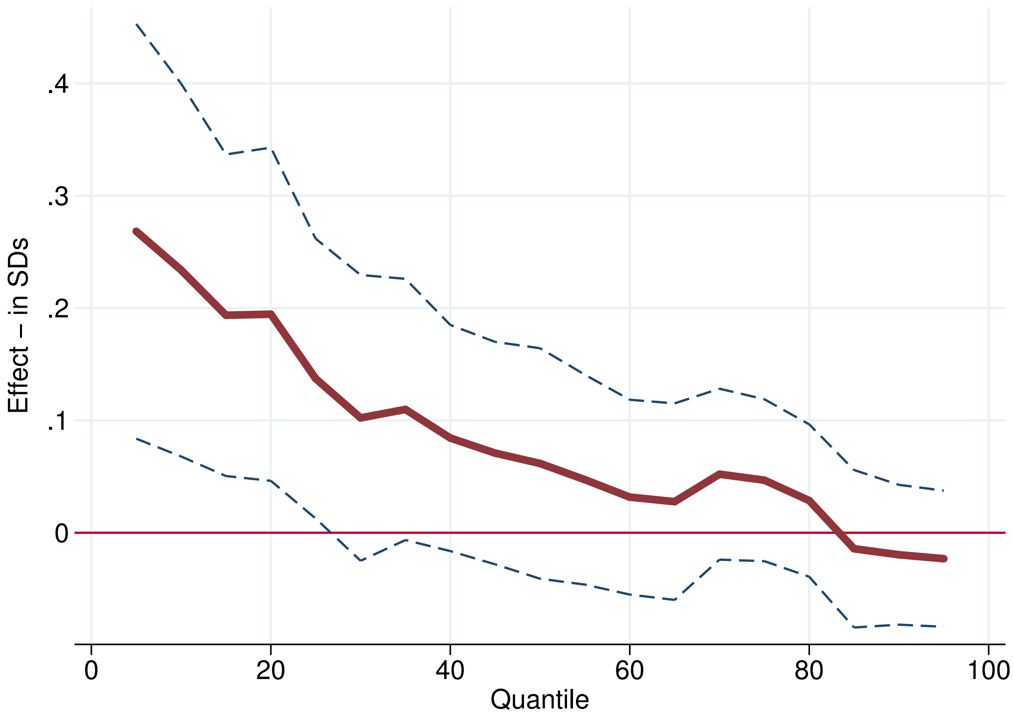
Figure 3: Distribution of math test scores by teacher-student gender pairing



Notes: This figure plots the distribution of students' scores on math midterm examinations by the four possible teacher-student gender pairings. A gaussian kernel was used to generate the density plots. Test scores are standardized within each grade, within a given school, so that ten points is one standard deviation and the mean is 70.

combined distribution of the test scores of students in other teacher-student gender pairings with a p-value of less than 0.001. To formalize these patterns, we present quantile regression results in Figure 4, estimating equation 2 without any of the independent variables related to low perceived ability and recovering coefficient estimates of β_3 and the corresponding confidence interval at every fifth centile between the fifth and 95th. Consistent with our visual inspection of Figure 3, the quantile regression results show that the gains from teacher-student gender match accrue to those girls in the left half of the distribution, with the largest gains accruing to those in the first quartile. Note that even the largest of these quantile estimates are smaller than the estimate of γ_3 in Table 4. We interpret this as further evidence of the contribution our model, which predicts the largest benefits among the low *perceived* ability children, as opposed to merely those children who place

Figure 4: Quantile regression results for math test scores



Note: This figure presents coefficient estimates and standard errors of β_3, estimated at every fifth quantile from the fifth to 95th, using equation 2 but removing the low perceived ability controls and their interactions (i.e., all of the terms with γ coefficients). The dependent variable is midterm math test score.

lower in the true ability distribution.

In column 4 of Tables A.5 and A.6, we estimate a smaller but still positive and significant effect of teacher-student gender match on math test scores for our alternative specifications of low perceived ability. While the below-median and quartile results are both sizable, the framework in Section 2 predicts that it is specifically among the low perceived ability girls, not just the low performers, that we should see the largest difference. Results from both specifications of low perceived ability bear out this prediction.

5.4 Persistence of effects

A year after the first wave of CEPS data was collected, the CEPS collected a second wave of data from the subset of children who were in the seventh grade during the first wave. This data includes perceived difficulty of current math class, time spent in math tutoring, job aspirations, and score on the standardized eighth grade midterm math test. This allows us to estimate the impact of teacher-student gender match, one year later, for this subset of students.

We estimate the impact of teacher-student gender match in grade 7 on these downstream outcomes, presenting results in Table 5. We find that while the estimated impact of female math teachers on low perceived ability girls' perceived difficulty of math and hours spent in tutoring disappears, the estimated effects on aspirations and test scores persist, and are of similar magnitude to estimated effects on these outcomes in the seventh grade.

Unfortunately, our sample size is heavily constrained - the second wave contains less than two thirds of the original sample. Furthermore, because of the small sample size of the low perceived ability group and the further splitting of the sample into male and female teachers in 8th grade, we are unable to use this data to precisely estimate the effects of having two female math teachers, or a male math teacher in the seventh grade and a female math teacher in the eighth.

6 Alternative explanations and limitations

In this section, we conduct a series of analyses to test for evidence of a series of alternative explanations for the results presented in the previous section. Our results suggest that the estimates we present in Section 5 are the result of the information sent by the teacher about each student's ability, and not the teacher's effort, teaching methods, or differential allocation of resources to students by gender. We then provide a discussion of the main limitations of our analysis.

Table 5: Persistence of effects of teacher-student gender match after one year

	(1) Perceived difficulty of current math class	(2) Aspires to jobs in art and design	(3) Hours per week spent in math tutoring	(4) Eighth grade midterm math test score
Girl x female teacher x low perceived ability	0.030	-0.116	1.051	0.421***
	(0.081)	(0.073)	(1.695)	(0.174)
Female teacher x low perceived ability	0.083	0.026	0.094	-0.228
	(0.060)	(0.041)	(1.253)	(0.149)
Girl x female teacher	-0.004	-0.002	-0.415	0.027
	(0.038)	(0.027)	(0.595)	(0.075)
Girl x low perceived ability	-0.010	0.110*	-1.013	0.025
	(0.058)	(0.062)	(1.206)	(0.130)
Girl	0.093***	0.236***	-0.047	0.181***
	(0.032)	(0.024)	(0.480)	(0.063)
Female teacher	-0.019	0.034**	-0.915	0.230***
	(0.030)	(0.016)	(0.639)	(0.081)
Low perceived ability	0.096**	-0.010	-0.577	-0.803***
	(0.048)	(0.032)	(0.999)	(0.111)
Mean for non-LPA boys	0.458	0.090	5.902	7.007
Number of observations	5,107	5,112	5,075	5,282

Note: this table shows estimated impacts of the impact of teacher-student gender match in the seventh grade on outcomes measured in the eighth grade. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 4 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

6.1 Alternative explanations

In this subsection we perform a series of analyses looking for evidence of alternative explanations for the patterns we observe. The intuition behind these tests is that the effects we estimate in the previous section could be driven by either teacher-specific characteristics or teacher conduct instead of by the effect of the teacher serving as a role model. We test for the following possibilities: one, that female math teachers give more attention to low perceived ability girls than do male teachers; two, that female math teachers are merely better teachers and it is these skill differentials which drive the observed effects; three, that our effect estimates are driven by female teachers exerting more effort than male teachers; and four, that our findings are driven by differences in teaching methods between female and male teachers.

First we investigate whether female teachers in our sample favor girls with more praise and attention (Beaman et al., 2009; Hoffmann and Oreopoulos, 2009; Jones and Wheatley, 1990). The CEPS collects students' recall of how frequently their current math teacher asks them questions and their recall of how frequently the teacher praises them in the classroom. In Table 6 we present results from estimating equation 2 using these two measures as outcome variables.[17] Our results show that while female teachers are slightly more likely to ask students questions than male teachers, there is no evidence that female teachers favor low perceived ability girls either with more opportunities to respond to questions or more praise.

The second possibility is that female teachers are simply better teachers, and it is teaching skill that drives the gains we observe for girls with low perceived ability. To test for this, we generate two sets of results. First, we replace the teacher-student gender match variable (i.e., girl x female math teacher) in our estimating equation with an interaction term for girl x math teacher who won an award. We show these results in Table 7. These results do not show any evidence of "better" teachers having a positive effect on perceived difficulty, aspirations, stereotypical beliefs, or performance of low perceived ability girls. To probe this further, we also conduct a horse race

[17] Responses are coded on a four-point scale, ranging from one for "strongly disagree" to four for "strongly agree." We break this into a binary variable, mapping strongly agree and somewhat agree to one, and somewhat disagree and strongly disagree to zero. The results we show are not sensitive to recoding the middle values in either direction.

Table 6: Robustness checks - teacher attention

	(1) Is called on frequently in math class	(2) Is praised frequently in math class
Girl x female teacher	0.035	-0.083
x low perceived ability	(0.080)	(0.069)
Female teacher	-0.084	-0.022
x low perceived ability	(0.055)	(0.043)
Girl x female teacher	0.008	0.030
	(0.025)	(0.024)
Girl	0.012	0.022
x low perceived ability	(0.053)	(0.048)
Girl	-0.030	-0.055***
	(0.022)	(0.020)
Female teacher	0.057	0.024
	(0.035)	(0.036)
Low perceived ability	-0.077*	-0.097***
	(0.043)	(0.030)
Mean for non-LPA boys	0.635	0.513
Number of observations	8,237	8,246

Notes: The regression specification used here is given in equation 2, again with the addition of the midterm math test score. For Column 1, the dependent variable is the response, on a four point scale from one, strongly disagree, to four, strongly agree, to the prompt "the teacher calls on me frequently." We code this as 0/1 for disagree/agree. Column 2's dependent variable, with the same scale and coding, 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. *p < 0.1, **p < 0.05, ***p < 0.01.

Table 7: Teacher aptitude

	(1) Perceived difficulty of current math class	(2) Aspires to jobs in art and design	(3) Midterm math test score
Girl x award-winning teacher x low perceived ability	0.040	0.105	-0.072
	(0.068)	(0.068)	(0.168)
Award-winning teacher x low perceived ability	-0.037	-0.018	0.157
	(0.055)	(0.035)	(0.134)
Girl x award-winning teacher	-0.012	-0.064***	0.049
	(0.015)	(0.019)	(0.052)
Girl x low perceived ability	-0.070	0.006	0.229*
	(0.049)	(0.042)	(0.133)
Girl	0.033***	0.219***	0.108***
	(0.010)	(0.015)	(0.037)
Award-winning teacher	0.005	0.055***	-0.081
	(0.021)	(0.019)	(0.092)
Low perceived ability	0.519***	-0.013	-0.956***
	(0.039)	(0.024)	(0.099)
Mean for non-LPA boys	0.122	0.104	7.024
Number of observations	8,276	8,213	8,294

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

regression, reverting to the original specification in equation 2 and adding a term to the right hand side interacting the award-winning teacher and the teacher-student gender match dummies. We find the interaction term is negative and insignificant, while γ_3 is of similar magnitude (e.g., for the test score results, 0.3 SD or larger) and retains its statistical significance.[18] We generate (but do not show) similar results for two separate specifications. The first swaps the receipt of a teaching award with years of experience; the second, with holding a degree from a teacher training (normal) university. These analyses similarly show no impact of teacher accolades affecting either low perceived ability student outcomes or any evidence that their inclusion in the regression changes the magnitude of our estimate of γ_3.

Next, we investigate the possibility that teacher effort drives our results. The CEPS collects self reported time use data from teachers. We use the following data points: first, how many hours teachers spend preparing for class and grading homework, respectively. We use these as proxies for how much "effort" the teacher chooses to expend. Second, how many hours the teacher spends lecturing. We use this as a scale variable - schools determine how many classes the teacher is responsible for, which is the denominator by which we scale our raw measure of effort. We generate three measures of effort: one, [hours in preparation: hours in class]; two, [hours grading: hours in class]; and three, [(hours in preparation + hours grading): hours in class]. We use these to estimate the effect of differential effort levels, between teachers within a given school, on student outcomes. We estimate the effects of effort in the same way we estimate the effects of teacher-student gender match, only now our independent variable of interest is the interaction of teacher effort and student gender. Because of the volume of results this generates, we describe the results here and present the results, in tabular form, in the appendix (Tables A.7, A.8, and A.9). Our results show that, for the same dependent variables - perceived difficulty of math, enrollment in math tutoring, and the math midterm exam score - we are unable to find a significant relationship between any of our teacher effort measures and child outcomes with the predicted pattern of effects among low perceived ability children.[19]

[18]For brevity, results in tabular form are available from the authors but not included in this manuscript.

[19]We cannot entirely exclude an alternative explanation for this pattern: teachers who expend less effort in terms of

Finally, we look at the impact of teachers' use of different methods of teaching on low perceived ability girls and boys. This tests for the possibility that the effects we observe are driven merely by female teachers employing different methods - e.g., engaging with students in a different way - which may affect low perceived ability girls and boys differentially. The CEPS records teachers' response to the following question: "how often do you use [teaching method]: never, sometimes, often, or always?" The question is asked separately for each of three methods, "lecturing," "small group discussion," and "interactive discussion between teacher and students." The latter two options involve more interaction between the student and teacher and so we expect, a priori, for them to have a larger effect on the low performing girls if teaching method does in fact drive the results in Section 5. As with the student engagement variables, there are four possible responses for how often teachers use these methods - never, sometimes, often, and always. We code these as a binary variable, with "often" and "always" mapping to one and the other responses to zero. Table A.10 shows estimates of the effect of teachers' use of these methods on perceived difficulty of math and midterm math test scores. We see no positive effect of using either method on low perceived ability girls' outcomes.

6.2 Limitations

In this subsection, we outline a few limitations of our analysis. First, this study looks at the effects of teacher-student gender match in mathematics, a subject where girls face longstanding stereotypes against their ability. A good ancillary test of our theory would be to test for effects of teacher-student gender match on beliefs and test performance in English and Chinese, subjects without stereotypes. Unfortunately, the very small number of male English or Chinese teachers in our data prevents us from using our identification strategy, which clusters at the grade-by-school level, to precisely test for such effects. We present the distribution of Chinese and English test scores by teacher-student gender pairing in Figure A.3, and note that the large difference seen in the math scores of low perceived ability girls assigned to female teachers is not apparent; overall,

out of class hours may merely be more productive with their time.

we see some evidence of (substantially smaller) effects of teacher-student gender match on beliefs and test scores in these two subjects. Whether because of small actual effect sizes, because of the large standard errors and the reduced sample size - we have to exclude grades in schools without at least one male Chinese or English teacher - we are unable to reject a zero effect.

Second, we observe a change in children's enrollment in tutoring which admits several possible explanations. One possible explanation at odds with our interpretation of the results is that parents' and/or teachers' compensatory actions, including but not limited to enrolling low perceived ability boys (girls) in less (more) tutoring, causes the changes in child outcomes. While this may play some role for some students, the patterns in our empirical results and a few facts about the Chinese context suggest this is unlikely to be the most important driver of our empirical results. First, our analyses of teacher effort and interaction with students by gender show no evidence of differential teacher attention or effort affecting low perceived ability children. Second, were parents' compensatory behavior to drive this pattern, parents of boys assigned to female teachers would be responding by withdrawing their children from tutoring while the parents of girls assigned to female teachers respond by increasing enrollment in tutoring. Our explanation is that these results come from a difference in children's enthusiasm, effort, and belief in themselves, generated by the role model effect of being assigned a same-gendered teacher for low perceived ability girls and, for low perceived ability boys, the identity threat of encountering a woman teaching subject in which stereotypes suggest men are superior. These explanations are rooted in existing empirical and theoretical evidence from both economics and psychology (e.g., Bettinger and Long, 2005; Nixon and Robinson, 1999; Paredes, 2014; Lybbert and Wydick, 2016b; Bian et al., 2017). Lastly, we study a context - Chinese middle schools - where existing evidence suggests children are often actively involved in their educational decisions. Loyalka et al. (2013) find that an information intervention providing students in a different set of Chinese middle schools with estimated labor market returns to different levels of education affected these students' propensity to drop out of middle school. This evidence is consistent with the notion that children in Chinese middle schools make at least some of their own educational decisions.

7 Conclusion

In this paper, we study how stereotypes and role models affect the formation of children's beliefs about their ability and how this affects their performance in mathematics. We model this process and generate a set of three predictions, which center around one key intuition: children who perceive themselves to be of low ability in math are the most likely to be affected by being assigned a female math teacher. We test these predictions using nationally representative data from Chinese middle schools, and our empirical results bear out the model's predictions. We find that low perceived ability girls benefit from being assigned a female math teacher, and low perceived ability boys are harmed by being assigned a female math teacher. Female math teachers appear to have little effect for students, regardless of gender, who do not perceive themselves to be of low ability.

Our paper generates two main messages. First, it shows that even in the increasingly common case of a reverse gender gap, i.e., where girls outperform boys, pro-male gender stereotypes may persist and harm the most vulnerable. This finding has clear implications for the assignment of students to teachers in subjects where such stereotypes persist. Second, it adds to a growing body of research studying how information affects aspirations, educational decisions, and outcomes (e.g., Bernard et al., 2014; Lusher et al., 2015; Lybbert and Wydick, 2016b; Genicot and Ray, 2017). As a whole, this work shows that the informational environment a child faces and, specifically, the presence of a plausible example of success, may be a key lever for changing beliefs, increasing effort, and improving performance in school, particularly for children who are uncertain of their own ability to succeed. More broadly, our study suggests that role models and the information they provide are likely an important input into the production of human capital; our results show that this is even more important among groups who, for various historical or socioeconomic reasons, may have their beliefs about their own ability artificially suppressed by exposure to certain types of information, such as stereotypes.

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.

Banerjee, Abhijit V. and Esther Duflo, "The economic lives of the poor," *Journal of Economic Perspectives*, 2007, *21* (1), 141.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, Petia Topalova et al., "Powerful Women: Does Exposure Reduce Bias?," *Quarterly Journal of Economics*, 2009, *124* (4), 1497–1540.

Becker, Gary S, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, National Bureau of Economic Research, Inc, 1975.

Bedard, Kelly and Stefanie Fischer, "Does the response to competition depend on perceived ability? Evidence from a classroom experiment," *Journal of Economic Behavior & Organization*, 2019, *159*, 146–166.

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," *BREAD Working Paper No. 429*, 2014.

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

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.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Beliefs about gender," *American Economic Review*, 2019, *109* (3), 739–73.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Burde, Dana and Leigh L Linden, "Bringing education to Afghan girls: A randomized controlled trial of village-based schools," *American Economic Journal: Applied Economics*, 2013, *5* (3), 27–40.

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

_ , Marianne E Page, and James E West, "Sex and science: How professor gender perpetuates the gender gap," *Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Chay, Kenneth Y, Patrick J McEwan, and Miguel Urquiola, "The central role of noise in evaluating interventions that use test scores to rank schools," *American Economic Review*, 2005, *95* (4), 1237–1258.

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.

Cunha, F. and J. Heckman, "The technology of skill formation," *American Economic Review*, 2007, *97* (2), 31–47.

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.

Eble, Alex and Feng Hu, "The sins of the parents: Persistence of gender bias across generations and the gender gap in math performance," *CDEP-CGEG Working Paper 53*, 2018.

Genicot, Garance and Debraj Ray, "Aspirations and Inequality," *Econometrica*, 2017, *85* (2), 489–519.

Gong, Jie, Yi Lu, and Hong Song, "The effect of teacher gender on students' academic and noncognitive outcomes," *Journal of Labor Economics*, 2018, *36*, 743–778.

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.

Hansen, Ben B and Jake Bowers, "Covariate balance in simple, stratified and clustered comparative studies," *Statistical Science*, 2008, pp. 219–236.

He, Leshui, Stephen L Ross et al., "Classroom peer effects and teachers: Evidence from quasi-random assignment in a Chinese middle school," *Human Capital and Economic Opportunity Global Working Group Working Paper 2017-014*, 2017.

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.

Hu, Feng, "Do girl peers improve your academic performance?," *Economics Letters*, 2015, *137*, 54–58.

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.

Jouini, Elyès, Paul Karehnke, and Clotilde Napp, "Stereotypes, underconfidence and decision-making with an application to gender and math," *Journal of Economic Behavior & Organization*, 2018, *148*, 34–45.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen, "Behavioral economics of education," *Journal of Economic Behavior & Organization*, 2015, *115*, 3–17.

Kofoed, Michael S et al., "The effect of same-gender and same-race role models on occupation choice: Evidence from randomly assigned mentors at West Point," *Journal of Human Resources*, 2017, pp. 0416–7838r1.

Latane, Bibb, "The psychology of social impact.," *American Psychologist*, 1981, *36* (4), 343.

Lavy, Victor and Analia Schlosser, "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 2011, *3* (2), 1–33.

Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland, "Expectations of brilliance underlie gender distributions across academic disciplines," *Science*, 2015, *347* (6219), 262–265.

Lim, Jaegeum and Jonathan Meer, "The impact of teacher-student gender matches: Random assignment evidence from South Korea," *Journal of Human Resources*, 2017, *52* (4), 979–997.

_ **and** _ , "Persistent Effects of Teacher-Student Gender Matches," *Journal of Human Resources*, Forthcoming.

Loyalka, Prashant, Chengfang Liu, Yingquan Song, Hongmei Yi, Xiaoting Huang, Jianguo Wei, Linxiu Zhang, Yaojiang Shi, James Chu, and Scott Rozelle, "Can information and counseling help students from poor rural areas go to high school? Evidence from China," *Journal of Comparative Economics*, 2013, *41* (4), 1012–1025.

Lusher, Lester, Doug Campbell, and Scott Carrell, "TAs like me: Racial interactions between graduate teaching assistants and undergraduates," *National Bureau of Economic Research*, 2015, (w21568).

Lybbert, Travis J and Bruce Wydick, "Hope as aspirations, agency, and pathways: Poverty dynamics and microfinance in Oaxaca, Mexico," *National Bureau of Economic Research Working Paper*, 2016, (w22661).

_ and _ , "Poverty, aspirations, and the economics of hope," *Economic Development and Cultural Change*, 2016.

Macours, Karen and Renos Vakis, "Changing households' investment behaviour through social interactions with local leaders: Evidence from a randomised transfer programme," *The Economic Journal*, 2014, *124* (576), 607–633.

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.

_ and Venkatesh Sundararaman, "Teacher performance pay: Experimental evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

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.

Porter, Catherine and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models," *Working paper*, 2017.

Rodríguez-Planas, Núria and Natalia Nollenberger, "Let the girls learn! It is not only about math...it's about gender social norms," *Economics of Education Review*, 2018, *62*, 230 – 253.

Ross, Phillip H, "Occupation Aspirations, Education Investment, and Cognitive Outcomes: Evidence from Indian Adolescents," *Working Paper*, 2017.

Scheepers, Daan and Naomi Ellemers, "When the pressure is up: The assessment of social identity threat in low and high status groups," *Journal of Experimental Social Psychology*, 2005, *41* (2), 192–200.

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.

Steele, C.M. and J. Aronson, "Stereotype threat and the intellectual test performance of African Americans.," *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, 1995, *69* (5), 797.

Steele, Jennifer, "Children's gender stereotypes about math: The role of stereotype stratification," *Journal of Applied Social Psychology*, 2003, *33* (12), 2587–2606.

Wilson, W.J., *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press, 2012.

Appendix - for online publication only

Appendix A: Appendix tables

Table A.1: Summary statistics for students

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

Note: This table uses only data from the main estimation sample in the paper, described in Section 4.1.

Table A.2: Summary statistics for teachers

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

Notes: This table compares observable teacher characteristics across teacher gender. This table also uses only data from the main estimation sample in the paper, described in Section 4.1.

Table A.3: Tests for gender-specific teacher quality

	(1)	(2)
Age	-0.010	-0.018
	0.009	0.030
Has B.A.	0.122	0.055
	0.078	0.249
Went to teachers' college	-0.242*	-0.222
	0.131	0.216
Years of experience	0.001	0.015
	0.008	0.027
Won award at province level	0.099	0.161
	0.115	0.387
Won award at city level	-0.027	-0.108
	0.073	0.255
Number of observations	207	207
R-squared	0.06	0.70
Joint test F-statistic	2.31	0.25
[p-value]	[0.04]	[0.96]
Grade-by-school fixed effects		X

Notes: This table shows coefficient and standard error estimates from regressing teacher gender on the predetermined teachers characteristics listed in the first column and conducting a Wald Test for their joint significance, similar to the results shown in Table 1 for student characteristics. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.4: Background characteristics, summarized by gender and perceived ability

	Perceived ability			
	Low		*Not low*	
	(1) Girls	(2) Boys	(3) Girls	(4) Boys
Age	13.50	13.52	13.13	13.26
Ethnic minority	0.23	0.19	0.11	0.10
Holds agricultural hukou	0.56	0.64	0.47	0.49
Number of siblings	1.06	0.93	0.72	0.63
Low household income / poor	0.30	0.30	0.16	0.19
Father's years of schooling	9.47	9.30	10.86	10.68
Mother's years of schooling	8.41	8.41	10.21	9.92
Number of observations	536	471	3,934	4,351

Notes: this table shows group-specific means for the low perceived ability girls and boys in our sample and, separately, for those who are not low perceived ability.

Table A.5: Replicating main results, using alternative definition of low perceived ability

	(1) Perceived difficulty of current math class	(2) Aspires to jobs in art and design	(3) Midterm math test score
Girl x female teacher x low perceived ability (alternate definition)	-0.113***	-0.053	0.274***
	(0.040)	(0.036)	(0.109)
Female teacher x LPA alternate definition	0.027	0.012	-0.110
	(0.032)	(0.018)	(0.086)
Girl x female teacher	-0.025	0.014	-0.039
	(0.016)	(0.022)	(0.061)
Girl x LPA alternate definition	0.078***	0.076***	-0.065
	(0.031)	(0.026)	(0.077)
Girl	0.021	0.169***	0.212***
	(0.013)	(0.019)	(0.053)
Female teacher	0.017	0.000	0.190***
	(0.018)	(0.018)	(0.067)
Low perceived ability (alternate definition)	0.180***	-0.015	-0.647***
	(0.025)	(0.014)	(0.064)
Mean for non-LPA boys	0.094	0.100	7.137
Number of observations	8,276	8,213	8,294

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.6: Replicating main results, using below median test score instead of perceived ability

	(1) Perceived difficulty of current math class	(2) Aspires to jobs in art and design	(3) Midterm math test score
Girl x female teacher x below median	-0.078***	0.009	0.111*
	(0.027)	(0.034)	(0.057)
Female teacher x below median	0.007	-0.004	0.087***
	(0.024)	(0.019)	(0.033)
Girl x female teacher	-0.039**	-0.013	0.012
	(0.017)	(0.022)	(0.029)
Girl x below median	0.068***	0.026	0.056
	(0.022)	(0.026)	(0.046)
Girl	0.058***	0.188***	0.047**
	(0.014)	(0.019)	(0.024)
Female teacher	0.026	0.006	0.061**
	(0.020)	(0.018)	(0.030)
Below median	-0.031	-0.002	-1.653***
	(0.021)	(0.023)	(0.029)
Mean for above median boys	0.069	0.085	7.757
Number of observations	8,300	8,251	8,345

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.7: Teacher effort 1

	(1) Perceived difficulty of current math class	(2) Enrolled in after-school math tutoring	(3) Midterm math test score
Hours prep: hours in class	0.037	0.021	-0.062
x low perceived ability x girl	(0.056)	(0.033)	(0.123)
Hours prep: hours in class	-0.018	-0.006	0.152**
x low perceived ability	(0.032)	(0.024)	(0.076)
Hours prep: hours in class	-0.007	-0.011	0.051
x girl	(0.007)	(0.012)	(0.032)
Girl x low perceived ability	-0.093	-0.030	0.283
	(0.073)	(0.055)	(0.177)
Girl	0.036***	0.051***	0.075
	(0.011)	(0.018)	(0.048)
Hours prep: hours in class	0.019*	-0.033***	-0.097*
	(0.011)	(0.010)	(0.058)
Low perceived ability	0.520***	0.003	-1.055***
	(0.046)	(0.031)	(0.108)
Mean for non-LPA boys	0.122	0.210	7.024
Number of observations	8,212	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.8: Teacher effort 2

	(1) Perceived difficulty of current math class	(2) Enrolled in after-school math tutoring	(3) Midterm math test score
Hours grading: hours in class x low perceived ability x girl	0.013	0.022	-0.099
	(0.042)	(0.050)	(0.130)
Hours grading: hours in class x low perceived ability	-0.053	0.000	0.149*
	(0.036)	(0.030)	(0.087)
Hours grading: hours in class x girl	-0.011	0.006	0.051*
	(0.007)	(0.012)	(0.030)
Girl x low perceived ability	-0.071	-0.031	0.319*
	(0.064)	(0.055)	(0.186)
Girl	0.041***	0.034**	0.075*
	(0.011)	(0.017)	(0.045)
Hours grading: hours in class	0.014	-0.004	0.075
	(0.013)	(0.016)	(0.057)
Low perceived ability	0.559***	-0.004	-1.050***
	(0.051)	(0.034)	(0.122)
Mean for non-LPA boys	0.122	0.210	7.024
Number of observations	8,212	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.9: Teacher effort 3

	(1) Perceived difficulty of current math class	(2) Enrolled in after-school math tutoring	(3) Midterm math test score
Hours prep + grading: hours in class x low perceived ability x girl	0.016	0.016	-0.053
	(0.032)	(0.022)	(0.076)
Hours prep + grading: hours in class x low perceived ability	-0.024	-0.002	0.105**
	(0.022)	(0.016)	(0.051)
Hours prep + grading: hours in class x girl	-0.006*	-0.001	0.033*
	(0.004)	(0.007)	(0.018)
Girl x low perceived ability	-0.090	-0.040	0.334
	(0.082)	(0.058)	(0.212)
Girl	0.042***	0.043**	0.057
	(0.012)	(0.019)	(0.050)
Hours prep + grading: hours in class	0.012*	-0.015*	-0.020
	(0.006)	(0.008)	(0.035)
Low perceived ability	0.554***	-0.000	-1.121***
	(0.058)	(0.036)	(0.134)
Mean for non-LPA boys	0.122	0.210	7.024
Number of observations	8,212	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 2 are coded as (0 = No , 1 = Yes). The test score results in column 3 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

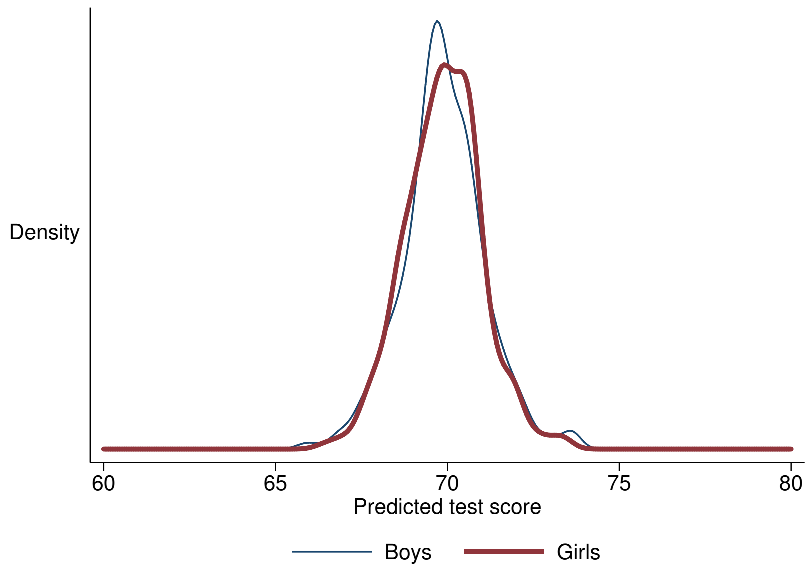
Table A.10: Teaching method

	Discuss in small groups		*Students and teacher "interactively" discuss*	
	(1) Perceived difficulty of current math class	(2) Midterm math test score	(3) Perceived difficulty of current math class	(4) Midterm math test score
Uses teaching method x girl x LPA	0.011	-0.103	0.048	-0.157
	(0.063)	(0.153)	(0.066)	(0.155)
Uses teaching method x LPA	0.028	0.001	-0.030	-0.115
	(0.048)	(0.123)	(0.050)	(0.131)
Uses teaching method x girl	-0.008	0.022	-0.033	-0.052
	(0.013)	(0.049)	(0.022)	(0.062)
Girl x LPA	-0.056	0.254**	-0.083*	0.310***
	(0.048)	(0.122)	(0.043)	(0.109)
Girl	0.031***	0.119***	0.053***	0.167***
	(0.011)	(0.033)	(0.020)	(0.054)
Uses teaching method	0.011	-0.046	0.007	-0.014
	(0.020)	(0.095)	(0.028)	(0.116)
Low perceived ability (LPA)	0.488***	-0.889***	0.520***	-0.801***
	(0.027)	(0.089)	(0.035)	(0.107)
Mean for non-LPA boys	0.122	7.024	0.122	7.024
Number of observations	8,257	8,275	8,251	8,268

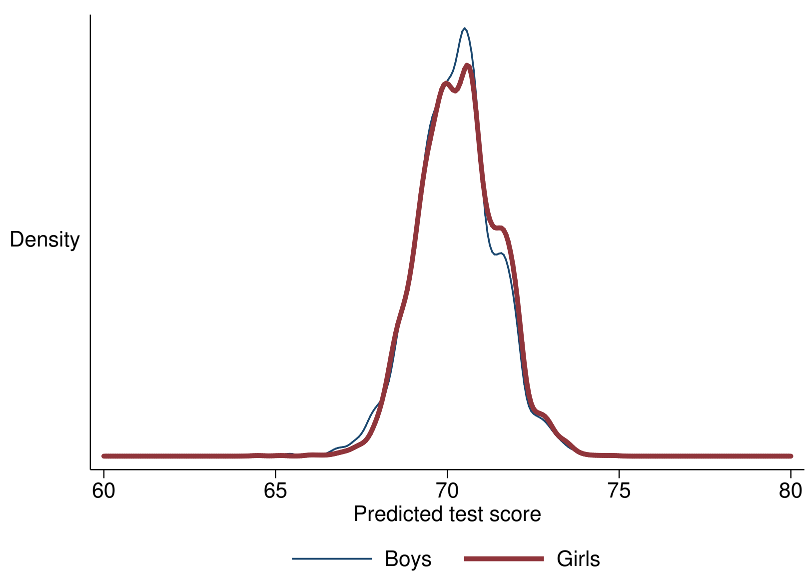
Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1 and 3 are coded as (0 = No , 1 = Yes). The test score results in columns 2 and 4 are presented in SD units. Robust standard errors clustered at the school level are shown in parentheses, and the coefficients are estimated using the specification in equation 2. *p < 0.1, **p < 0.05, ***p < 0.01.

Appendix B: Appendix figures

Figure A.1: Predicted test score distributions, by perceived ability



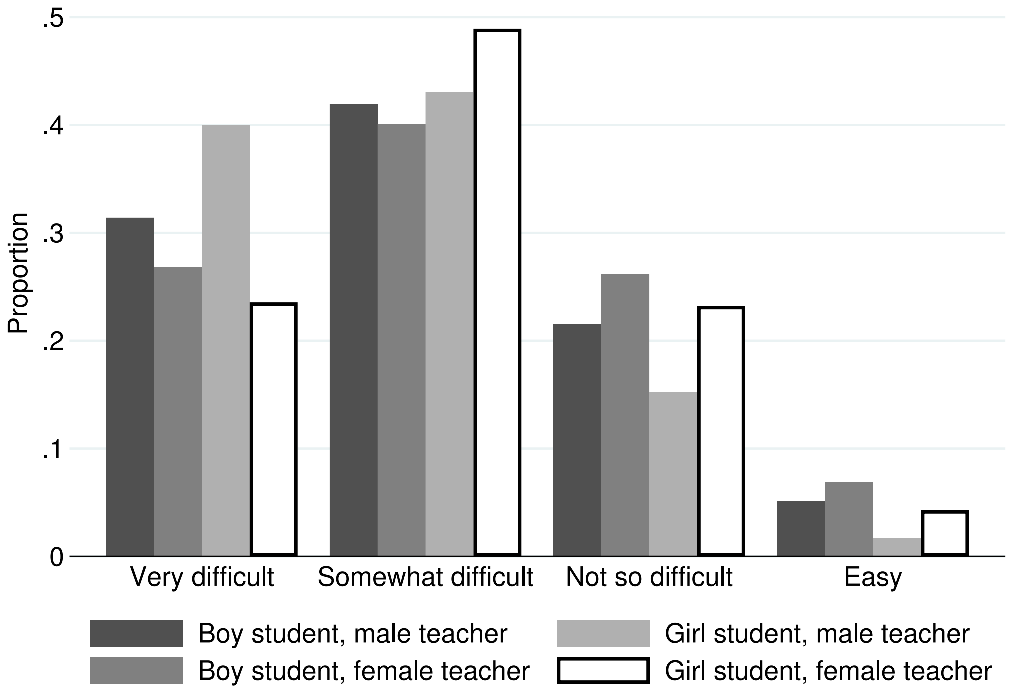
Panel A: Low perceived ability



Panel B: Not low perceived ability

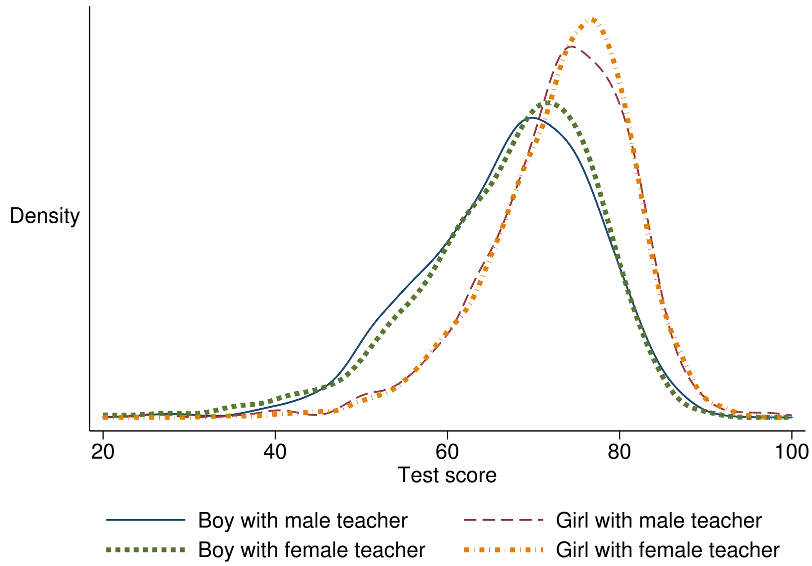
Notes: to generate these figures, we regress test scores on the vector of student-level predetermined characteristics and, using these coefficients, generate a predicted test score for each student. We then plot these using a gaussian kernel for each perceived ability-gender group.

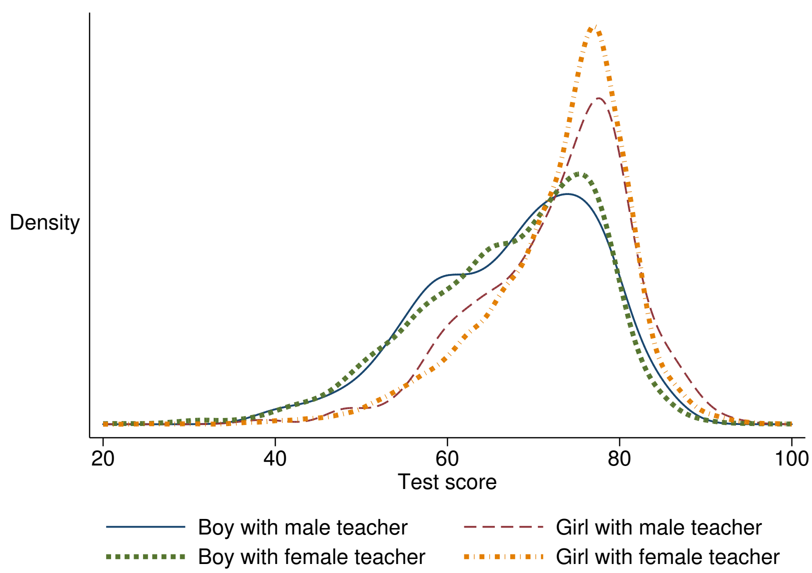
Figure A.2: Effect of teacher-student gender match on student beliefs, for those below within-group median test score



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.

Figure A.3: Distribution of English and Chinese test scores by teacher-student gender pairing





Notes: this figure shows the analogue to Figure 3 for English and Chinese scores. Note that the large difference between girls assigned to female teachers and all other pairings does not appear hear to the same extent that it does for math scores.

Appendix C: A more formal version of our conceptual framework

To capture the discrete nature of investment behavior below and above $\underline{r}$, and for ease of exposition, we divide beliefs about r^i into a discrete variable A_i^t, the belief of individual i about her ability at time $t \in \{1,2\}$:

$$A_t^i = \begin{cases} L & if \;\; r^i < \underline{r} \\ H & if \;\; r^i \geq \underline{r} \end{cases} \tag{3}$$

Our object of interest is a set of conditional probabilities $P(A_2^i = H|G^i, A_1^i, T^i)$, where the conditions relate to the gender of the student, her/his perceived ability in mathematics at time $t = 1$, and the gender of the middle school math teacher. We define student gender as $G^i \in \{G^f, G^m\}$, where f and m indicate the student is female or male, respectively. We define teacher's gender as $T^i \in \{T^f, T^m\}$, where the superscript again indicates gender. In the data we see that $P(A_t^i = H|G^f) < P(A_t^i = H|G^m)$, that is, girls have lower perceived ability in math than boys. Specifically, despite performing better on math tests than boys, girls are nearly 10 percentage points more likely to report that they find math at least somewhat difficult.[20] Furthermore, 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% vs. 39%).

Along with these data, we make four simple assumptions that allow us to generate the three predictions we test in our data.

Assumption 1: Given that girls have slightly better math test scores than boys, we assume that at least some of the gap in perceived math ability by gender is due to exposure to negative gender stereotypes.

Assumption 2: all teachers attempt to send the message that, with enough investment, a student can succeed in math, i.e., their goal is to convince students that $P(A_t^i = H)$. Note that in this

[20]Girls outperform boys in all subjects in our data, but this gap is smallest in math. In a separate paper (Eble and Hu, 2018), we study the contributors to this pattern and the tension between gendered beliefs about ability and actual performance in mathematics by gender.

framework, $P(A_t^i = H)$ signifies only that it is in the student's best interest to invest in her human capital, not that she is exceptionally gifted.

Assumption 3: $P(A_t^i = H|G^x, T^x) > P(A_t^i = H|G^x, T^{\sim x})$ for $x \in \{f, m\}$, that is, encountering a same-gendered math teacher delivers a more credible signal that the individual is likely to be H than a teacher of the other gender. We base this on the notion that a message sent by a teacher who shares an identity with the student provides a higher signal to noise ratio than that sent by a teacher with whom the student has no shared identity.[21]

Assumption 4: $P(A_t^i = H|G^f) > \zeta$, where ζ is some probability strictly greater than zero, ensuring that girls do not perceive themselves to be so unlikely to be of high ability that they will not update in response to a signal (that is, they are not in the leftmost portion of Figure A.4). From this, and the empirical fact that boys in our sample are on average more likely to perceive themselves as high ability in math, we get $E(|r^i - \underline{r}| \mid A_1^i = L) < E(|r^i - \underline{r}| \mid A_1^i = H)$. In prose: that those classified as low perceived ability are, in expectation, closer to the margin of deciding whether or not to invest in human capital than those not classified as low perceived ability.

Prediction 1: $P(A_2^i = H|G^f, T^f) - P(A_1^i = H|G^f) > P(A_2^i = H|G^m, T^z) - P(A_1^i = H|G^m)$, where $z \in \{f, m\}$. In prose, we predict that girls assigned to a female math teacher should update their prior on their ability to productively invest more than boys assigned to either a female or male math teacher. This is a direct consequent of the part of girls' lower perceived ability in math due to exposure to stereotypes, Assumption 4, and Bayes' rule.

Prediction 2:

$$P(A_2^i = H|G^f, A_1^i = L, T^f) - P(A_2^i = H|G^f, A_1^i = L, T^m) >$$
$$P(A_2^i = H|G^f, A_1^i = H, T^f) - P(A_2^i = H|G^f, A_1^i = H, T^m)$$

This prediction states that low-perceived ability girls will make larger (that is, more positive) updates to their prior than high perceived ability girls in response to encountering a female math teacher, as opposed to a male one. This prediction comes from Assumptions 3 and 4, and is de-

[21]These predictions are also derived in work on the psychological concept of "Social Impact Theory" (Latane, 1981).

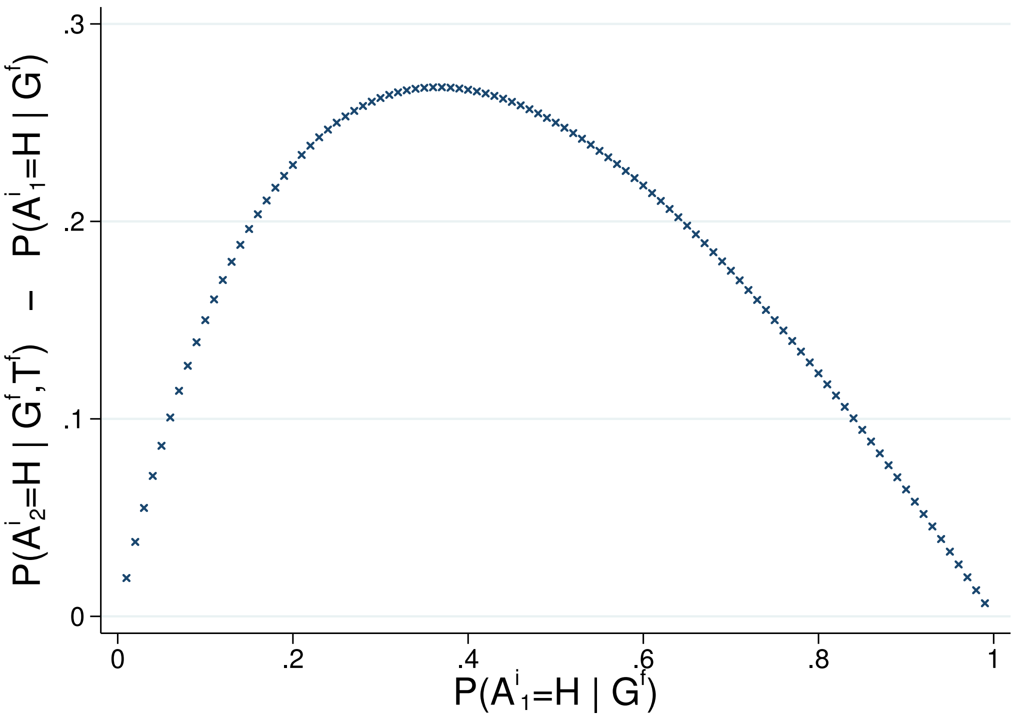
rived from a basic tenet of information theory: information that is relatively new to the receiver generates a larger update to the prior than it would among receivers for whom the information is less novel.[22] Seen through the lens of Bayesian updating, high perceived ability girls have a much higher $P(A_1^i = H)$ than low perceived ability girls. As a result, the same information causes much smaller updates for high perceived ability girls than for low perceived ability girls; Figure A.4 shows this result.

Corollary: depending on the proximity of r^i to $\underline{r}$, we should also see increases in s^i and academic performance among the low perceived ability girls assigned to female math teachers.

Prediction 3: $P(A_2^i = H | G^m, A_1^i = L, T^f) < P(A_2^i = H | G^m, A_1^i = L, T^m)$, that is, being assigned a female math teacher, as opposed to a male one, will reduce low perceived ability boys' belief in their ability to productively invest in themselves. This is derived from Assumption 1, that some of boys' greater confidence comes from exposure to gender stereotypes and so evidence in contradiction of those stereotypes will cause them to revise their beliefs downward. It can also be derived from the psychological concept of identity threat, which refers to the negative response (low performance, reduced effort) that occurs when members of a privileged group see a threat to the status quo (Scheepers and Ellemers, 2005). In our context, the existing stereotype posits that boys are better at learning math than girls. Low perceived ability boys, confronted with the dual threats of an increase in the difficulty of math when they enter middle school (described in the next section) and the appearance of a female math teacher, may interpret these as signals that threaten their perception of the status quo that, as boys, they are better than girls in learning math.

[22] That is, low perceived ability girls exposed to an example of success see it as more novel than do high perceived ability girls, who in themselves already have an example of success.

Figure A.4: Mapping of prior to size of update



Notes: this figure shows the mapping from a girl's prior that she is of high ability, $P(A_t^i = H|G^f)$, to the update of that prior in response to encountering a female math teacher. The assumptions used to generate this figure are $P(G^f, T^f|A_t^i = H) = 0.6$ and $P(G^f, T^f|A_t^i = L) = 0.2$, but the right-skewness of the mapping holds more generally under $P(G^f, T^f|A_t^i = H) > P(G^f, T^f|A_t^i = L)$.

Appendix D: Description of balanced assignment rule

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 in an alternating way - for the first five students, student 1 is assigned to class 1, student 2 is assigned to class 2, and so on until student 5. Then, student 6 is assigned to class 5, student 7 to class 4, an on until student 10 is assigned to class 1. Then the original order repeats, so that student 11 is assigned to class 1, student 12 to class 2, and on until student 15. At student 16, the order once again reverses, and so on, so as to avoid bifurcation of classrooms (that is, avoiding the case where the best and worst students are placed together in some classrooms and mid-level performers are placed together in others). This is described nicely in He et al. (2017), who, along with Hu (2015) and Gong et al. (2018), also exploit this quasi-random assignment of students to classes in Chinese middle schools.