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

December 2018

COLUMBIA | SIPA
Center on Global Economic Governance

Stereotypes, Role Models, and the Formation of Beliefs

Alex Eble and Feng Hu*

December 2018

Abstract

Information affects beliefs, which in turn determine investment decisions. Because the productivity of investment in human capital is partly determined by levels of prior investments, early sources of information play a crucial role in human capital formation. We study how information from stereotypes and role models influences children's beliefs, aspirations, investment, and academic performance. We use a simple, stylized model of investment under uncertainty to formalize how same-gendered math teachers may serve as role models for students exposed to stereotypes about gender and math ability, and how this affects the beliefs, behavior, and performance of different types of students. We exploit random assignment of students to classes in nationally-representative data from Chinese middle schools to test the main predictions of the model and address potential alternative explanations. Our empirical results are consistent with the model's predictions: being assigned a female math teacher generates large gains in beliefs, aspirations, investment, and test scores for low perceived ability girls, generates moderate harms for low perceived ability boys, and has no gender-specific

*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; O15.

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.

1 Introduction

As a person goes through life, her beliefs are shaped by what she comes in contact with. These beliefs inform her investment decisions and, more broadly, her aspirations for the future. A long line of both empirical and theoretical work has explored the importance of this causal chain in the formation of human capital (Becker, 1975; Akerlof and Kranton, 2002; Jensen, 2010; Lybbert and Wydick, 2016b; Genicot and Ray, 2017). Because the productivity of later human capital investments is affected by the levels of earlier investments (Cunha and Heckman, 2007), this link from information to beliefs and on to human capital investment decisions is particularly important in the early stages of life.

A crucial implication of this relationship is that early misinformation about ability can have long-lasting negative consequences. If telling a child she is of low ability in a given subject reduces her relevant investment in period one, in period two she will be relatively less able in the subject, bearing out the prediction of the misinformation and reinforcing its message. A common source of this type of misinformation is negative stereotypes about ability by gender and ethnicity (Steele and Aronson, 1995; Steele, 2003). Recent evidence suggests that gender stereotypes affect the interests and time use decisions of both girls and boys as early as age seven (Bian et al., 2017) and that these stereotypes may lead to underrepresentation of women and minorities in several scientific fields where such misinformation persists (Leslie et al., 2015).

In this paper, we study how children's beliefs are affected by stereotypes and role models, and how this influences aspirations, investment in skills, and academic performance. We first develop a simple stylized model based on that of Genicot and Ray (2017). In our model, children are uncertain about their own ability and the returns to exerting effort in school. They update their beliefs, as Bayesians, in response to the information they encounter. Stereotypes provide erroneous group-specific signals about ability. Role models also send signals about children's ability and, if the role model shares an identity with the child, the signal to noise ratio of her message is higher. Teachers serve as role models because they have frequent contact with students and occupy a place of authority and expertise (Latane, 1981; Bettinger and Long, 2005). Our model predicts that the greatest benefits of encountering a same-gendered teacher accrue to children facing a negative stereotype (girls) whose beliefs about their own ability place them near the margin of deciding whether or not to invest in human capital. It also predicts harms for lower-ability children whose beliefs about their own ability have been artificially inflated by exposure to the stereotype that those in their group are exceptionally good at a subject (boys) when these chilren are assigned to a teacher in that subject who is not of their gender.

We then take the predictions of our model to data from a nationally representative set of Chinese middle schools. In this setting, there is widespread belief among children that boys are better than girls in learning math. Due to a revision of China's compulsory education laws in 2006 which banned the tracking of entering middle school students to different classes based on academic performance, there is also random assignment of students to classrooms. This allows us to estimate the causal effects of different classroom configurations, a method one of us has used in previous work to study peer effects (Hu, 2015) and which has subsequently been used in other studies (e.g., He et al. 2017; Gong et al. 2018).

We estimate the effects of teacher-student gender match on girls' and boys' beliefs, behaviors, and academic outcomes, generating subgroup-specific estimates by whether the student perceives her/himself to be of low ability in math. In line with the predictions of our model, we find three sets of results. First, low perceived ability girls assigned to female math teachers are 20 percentage points less likely to perceive math 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, low perceived ability boys assigned to female math teachers are 10 perentage 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 estimate much smaller, non-significant effects of teacher-student gender match for children who are not of low perceived ability.

This difference between low perceived ability students and non-low perceived ability students highlights a key contribution of the paper. Our results are robust to two alternative specifications for defining low perceived ability, and persist when a subset of the students are tested again a year later. While the low perceived ability girl effect estimate is larger than commonly seen in the literature, the estimated effect of teacher-student gender match on *all* girls' test scores is 0.09 SD, well within the range of prior estimates (between 0.05 and 0.20 SD, c.f. Dee, 2007; Muralidharan and Sundararaman, 2011; Lim and Meer, Forthcoming). This suggests that, in our context, the phenomenon of teacher-student gender match improving girls' performance flows entirely to low percevied ability children. The harms for boys we observe are consistent with recent work on the "reverse gender gap" (Fortin et al., 2015) and the psychological concept of "identity threat" (Steele et al., 2002; Sherman et al., 2013). We argue that this link between our model and empirics advances understanding of one way teacher-student gender match may work, and for whom.

We then conduct a series of analyses to disentangle the different mechanisms which may drive the effects we observe. Our results suggest that the main (though by no means only) mechanism at work is female math teachers serving as role models for low perceived ability girls, protecting against the harmful effects of negative gender stereotypes. We provide three sets of analyses to support this claim. First, as described above, our model's empirical predictions are borne out in our data. We see large effects of being assigned a female teacher on the beliefs, aspirations, and investment behavior of low perceived ability girls, negative effects on low perceived ability boys, and no gender-specific effects for students - boys or girls - who are not low perceived ability. Second, 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, and find no evidence that any of these differentially affect low perceived ability students. Third, we test for the possibility that female teachers give extra attention, either praise or opportunity to speak in class, to low perceived ability girls, and find no evidence of such differential treatment.

Our work builds on two active lines of inquiry in economics. 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. Several recent papers, both empirical (e.g., Bernard et al., 2014; Macours and Vakis, 2014; Lybbert and Wydick, 2016a; Ross, 2016; Kofoed et al., 2017) and theoretical (e.g., Akerlof and Kranton, 2000, 2002; Bénabou and Tirole, 2011; Lybbert and Wydick, 2016b), have studied the role of aspirations in affecting investment behavior. We further this work by adding evidence of how role models - an important informational channel through which beliefs and aspirations can be influenced - may benefit children who face stereotypes or other sources of misinformation. Our results are consistent with both the hypothesis of Wilson (2012) and a key prediction of the model in Genicot and Ray (2017); namely, that informational shocks about oneself (as opposed to about the world, as in Jensen, 2010) may induce changes in aspirations, which in turn lead to changes in investment and outcomes.

The second is the set of studies estimating the effects of teacher-student identity match on the performance of stereotyped-against individuals (e.g., Dee, 2004; Bettinger and Long, 2005; Carrell et al., 2010; Gershenson et al., 2016; Lim and Meer, Forthcoming). Our analysis furthers understanding of one channel through which the benefits of match accrue, and for whom. The literature has established many ways in which a female math teacher can aid girls' learning, including but not limited to the role model mechanism. The correspondence between the predictions of our model and our empirical results, along with the ancillary evidence we present, together strongly suggest that the role model channel plays the dominant role in driving the changes in beliefs and performance that we measure for low perceived ability girls. This complements recent work showing that shared-identity teachers may serve as role models, 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 this paper is structured as follows. In Section 2 we lay out our conceptual framework and conduct a preliminary analysis of the test score data to illustrate the framework's main intuition. Section 3 describes the setting we study. Section 4.1 outlines our data sources and provides summary statistics of our main variables. Section 4.2 introduces our empirical strategy and presents results for tests of our main identifying assumptions. Section 5 presents our main empirical results estimating the effects of being assigned a female math teacher on student beliefs, aspirations, investment, and performance. Section 6 investigates possible mechanisms for these effects and discusses the limitations of our study. The final section concludes.

2 Conceptual framework

In this section, we introduce a simple framework of children updating their beliefs, as Bayesians, in the face of stereotypes. We overlay this on a canonical two-period consumption model with savings[1]. We use this framework to fix ideas about our main research question: how and for whom do stereotypes and role models affect the formation of beliefs about one's own ability? The model generates predictions we test later in the paper. We then justify a key assumption of the model and illustrate its main intuition with two simple analyses of our beliefs and test score data.

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 (Beilock et al., 2010; OECD, 2015). 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 beliefs about the returns to investment, causing girls to invest less effort, enthusiasm, and time in studying for math (Bian et al., 2017). This would lead to lower performance and thus a confirmation of the

[1] Our model draws heavily on the working paper version of Genicot and Ray (2017). It was initially conceived during Eble's PhD research. Its genesis predates the empirical work here by several years and informed our plan of analysis.

once-erroneous content of the stereotype. A corollary of these findings is that the presence of a female teacher in the same subject as the stereotypical belief could change girls' views about the potential positive returns to their effort in math. By virtue of shared gender, the female teacher provides girls with 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). Evidence from both psychology and economics also suggests that such an example could lead to an increase in students' academic motivation and expectations (Nixon and Robinson, 1999; Gershenson et al., 2016).

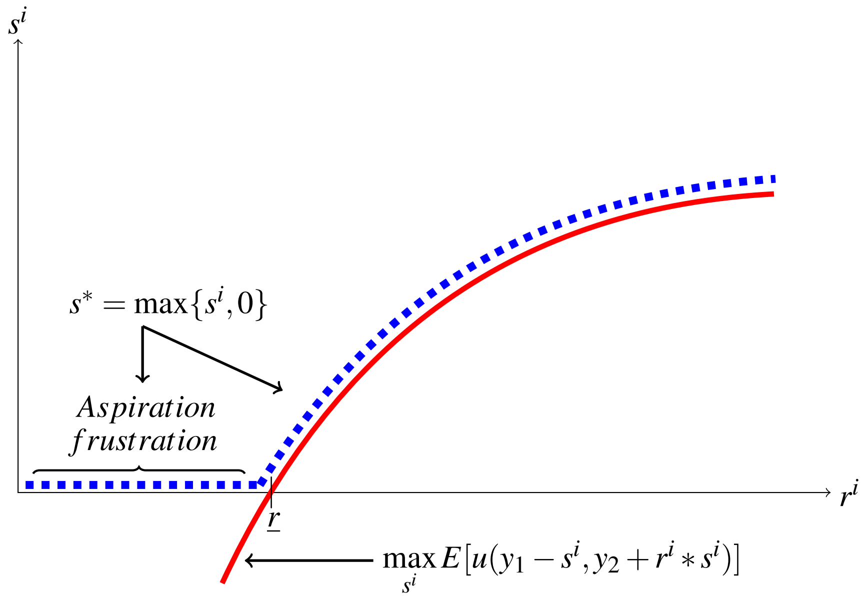
To formalize this intuition, we place our analysis in the context of a canonical two period model of consumption and savings. Individuals face the following consumption problem:

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

We index the individual in the superscript and time in the subscript. We assume utility from consumption is concave in both periods. Defining terms, y_t is income in period t, s^i is the savings of individual i in period 1, and r^i is 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, normalized 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, societal stereotypes, and so forth.

We introduce stereotypes and link to the Genicot and Ray (2017) model through r^i. We assume there is some 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

Figure 1: Visual depiction of the model



systematic biases such as stereotypes. These individuals will rationally but sub-optimally choose not to invest, part of what Genicot and Ray call "aspiration frustration." In Figure 1, we depict this static part of our model graphically, with s^* indicating the solution to the individual's optimization problem.

The key contribution of our conceptual framework is to model the evolution of beliefs about r^i over time. We model students as Bayesians, updating their beliefs about r^i in response to new information as it is received. The effect of new information on the update to r^i will depend on two parameters of the signal, its credibility and the difference between the individual's prior and the new information provided by the signal. Informally, for girl students who perceive themselves to be of sufficiently low ability in math that they fall into the aspiration frustration part of the support of r^i, being assigned a female 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 girls' priors, which have been formed both by gender stereotypes and the low signals about ability received up to that point. For other students, the combination of positive signals previously received (girls and boys who are doing well in math) and a lack of negative stereotypes (all boys) lead to the prediction that being assigned a same-gendered math teacher is unlikely to cause these individuals to update r^i in response to a female teacher as dramatically as will low perceived ability girls.

Formally, our Bayesian individual i proceeds through life gaining new information about r^i from her environment and experiences and updating her beliefs accordingly. 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 $t \in \{1,2\}$:

$$A_t^i = \begin{cases} L & if \ \ r^i < \underline{r} \\ H & if \ \ r^i \geq \underline{r} \end{cases} \tag{2}$$

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, in both periods in which we observe them, girls have lower perceived ability in math than boys[2]. We make three further assumptions that allow us to generate three predictions to test in our data.

Assumption 1: 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 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 2: $P(G^x, T^x|A_t^i = L) < P(G^x, T^x|A_t^i = H)$ for $x \in \{f, m\}$, that is, in either period, encountering a same-gendered math teacher delivers a signal that the individual is more likely to be H than L, both for boys and girls. We base this on the notion that teachers' role of authority gives their message some credibility and that a message sent by a teacher who shares an identity with the

[2]Given that girls have slightly better math test scores than boys, we argue that at least some of the gap in perceived ability is due to negative gender stereotypes.

student provides a higher signal to noise ratio than that sent by a teacher with whom the student has no shared identity[3]. This generates a mapping from $P(A_1^i = H|G^f)$ to $[P(A_2^i = H|G^f, T^f) - P(A_1^i = H|G^f)]$ that has a right-skewed inverse-U shape. We show an example of this in Figure A.1.

Assumption 3: $P(A_t^i = H|G^f) > \zeta$, 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.1).

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 girls' lower perceived ability, Assumption 3, 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 2 and 3, and is derived 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[4]. 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.1 shows this result.

Corollary: depending on the proximity of r^i to $\underline{r}$, we should also see increases in s^i and aca-

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

[4] 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.

demic performance among the low perceived ability girls assigned to female math teachers.

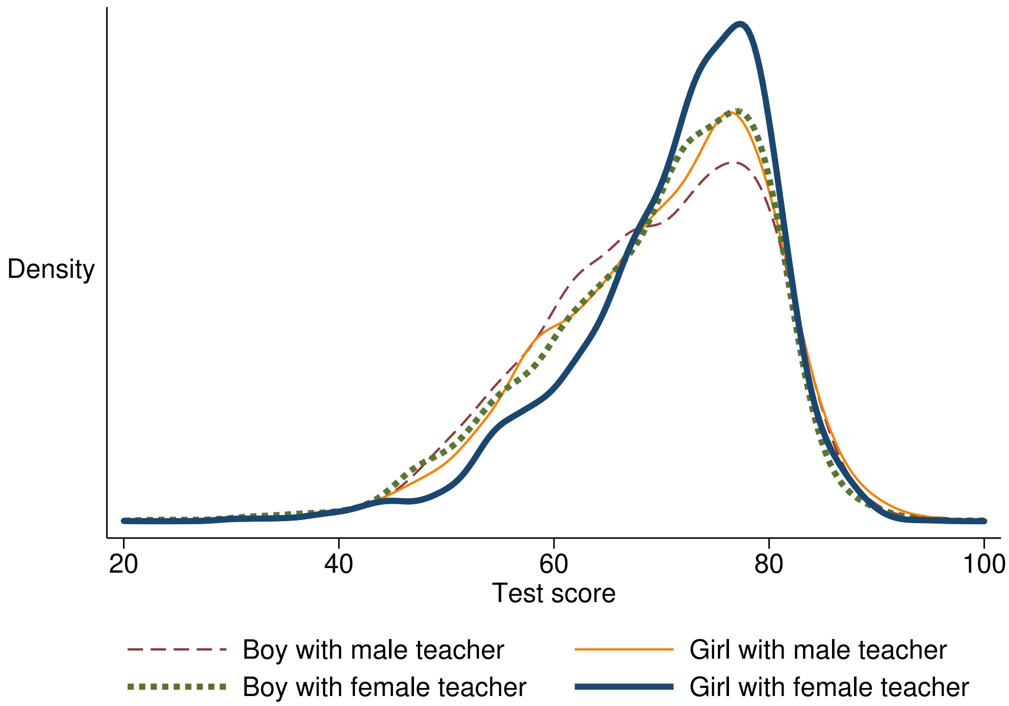
Prediction 3: $P(A_2^i = H|G^m, A_1^i = L, T^f) < P(A_2^i = H|G^m, A_1^i = L, T^m)$, that is, being assigned a female math teacher, as opposed to a male one, will reduce low perceived ability boys' belief in their ability to productively invest in themselves. This is derived from the psychological concept 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.

Next, we examine two features of our data - the distribution of perceived ability by gender and the distribution of math test scores - as evidence of the influence of stereotypes and as a preliminary test of our main prediction, respectively. Despite performing no worse on math tests than boys, girls are nearly 10 percentage points more likely to report that they find math at least somewhat difficult[5]. 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%). We offer this as prima facie evidence that negative gender stereotypes affect girls in our data, particularly low perceived ability girls.

Next, in Figure 2, 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 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 combined distribution of the test scores of students in other teacher-student gender pairings with a p-value of less than 0.001, and, as we will discuss further in Section 5, quantile regressions show substantial gains

[5]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.

Figure 2: 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. The sample is restricted to the estimation sample as described in Section 4. A gaussian kernel was used to generate the density plots. A Kolmogorov-Smirnov test rejects equality of the distributions of test scores between two groups: girls paired with a female teacher and the combined distribution of students in all other teacher-student gender configurations. Test scores are standardized within each grade, within a given school, so that ten points is one standard deviation and the mean is 70.

in the first through third deciles. As well as being in line with our main theoretical predictions, these results also suggest that we should look among students in the left tail of the (perceived) ability distribution for the potential impacts of teacher-student gender match.

3 Setting

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

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[6]. In the second system, students are assigned to classes by an algorithm which takes into account their academic performance at the beginning of the seventh grade and enforces a "balanced assignment" rule, requiring that the average quality of students be comparable across classes and the class not be bifurcated (Carman and Zhang, 2012). 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). In Appendix C, we provide a description of this type of assignment rule, borrowing from He et al. (2017). 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 work, we will exploit these two methods of assigning students to classes as providing potentially quasi-random matching of student gender to teacher gender. We note, however, that this system is not implemented with perfect fidelity, particularly as students move beyond

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

the first grade of middle school, i.e., from the seventh grade to the eighth and ninth. Unlike in many western countries, where admission to high school or university is either according to residence or based on multiple dimensions (e.g., grades and teacher recommendations), China's high school admissions system relies almost exclusively on entrance examination scores (Zhang, 2016). Furthermore, the promotion of middle school administrators is largely determined by their school's students' performance on the high school entrance examination. More specifically, promotion is often awarded according to the annual number of graduates admitted to elite high schools. As a result, despite the banning of class tracking some middle schools assign students to classes based on their academic performance in order to better prepare top students for the entrance examination. Along with this sorting, school administrators may channel better teachers and more resources to classes with higher-ability students to maximize the chances that some of these students place in the best high schools. This practice is more common in the eighth and ninth grades than in the seventh and means that after their first semester or year of middle school, students may be reassigned to different classes based on their academic performance even if they are randomly assigned at the beginning of the seventh grade. In this analysis, as in Hu (2015) and Gong et al. (2018), we restrict our attention to students randomly assigned to classes in the 7th grade and 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 same data and identification strategy, but analyze both a different subsample of the data and a different set of research questions. Gong et al. 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). Our paper studies how these match effects work in the face of gender stereotypes and, guided by the model, focuses on the effects for low perceived ability girls. To achieve this, we restrict our sample to mathematics classes, the only subject where anti-girl stereotypes prevail, and decompose effects by the student's perception of her own ability. Gong et al. instead study all children and do not differentiate between low and high perceived ability groups or subjects which do or do not have anti-girl stereotypes. Using their sample (all classes, not just mathematics) and removing the perceived ability interaction terms, we are able to reproduce their findings.

4 Data and empirical strategy

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

4.1 Data sources

The main data source 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[7]. The CEPS is a nationally representative longitudinal survey that aims to track middle school students through their educational progress and later labor market activities. The baseline survey of the CEPS adopted a stratified, multistage sampling design with probability proportional to size, randomly selecting approximately 20,000 seventh and ninth grade students from 438 classes in 112 schools from 28 counties across mainland China during the 2013-2014 academic year. 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) 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 and their families, as well as detailed information on students' beliefs, aspirations, and time use. It also collects administrative school

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

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 8, 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 biased for low perceived ability girls assigned to female teachers[8].

The teacher questionnaire contains rich information on teacher characteristics, including teachers' age, gender, education levels, years of teaching experience, whether the teacher graduated from a university for teachers, whether the teacher holds a senior professional rank, and whether the teacher has won any teaching awards at various levels. The survey also contains information on the subject and the class the teacher taught during the 2013-2014 academic year. We limit most of our analyses to the matched math teacher-student dataset.

The survey also collects data on the assignment mechanism used to assign students to classrooms, collected both from school principals and homeroom teachers[9]. 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 en-

[8]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.

[9]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.

tered the eighth or ninth grade. 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[10].

Table A.1 presents summary statistics for students by gender for those students randomly assigned to classrooms. 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.

Table A.2 shows summary statistics for teachers in the classrooms studied in Table A.1. In our data, 39% of the students are taught by male math teachers, alleviating the challenge faced in Antecol et al. (2015) where there was an insufficient number of male teachers to draw strong conclusions from some of the comparisons made. Female math teachers are on average younger and less experienced than their male counterparts. However, female teachers appear to be more qualified than their male counterparts in terms of education and proportion having won a teaching award at the province or national level[11].

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 stemming from such heterogeneity between teachers, between schools. Specifically, our empir-

[10]This is assumption also investigated in Hu (2015).

[11]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.

ical strategy compares male and female teachers within a grade, within a school; the observed differences attenuate dramatically and cease to be significant at this level of comparison.

4.2 Empirical strategy

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

In this paper we exploit the random assignment of students to classrooms to estimate the impact of teacher-student gender match on students' beliefs, aspirations, investment behavior, and performance on in-school examinations. We estimate a reduced form regression equation, controlling for grade-by-school fixed effects and vectors of observable, predetermined characteristics at the child and teacher levels. Specifically, to determine whether teacher-student gender match differentially affects the gender gap in outcomes of interest, we estimate the following equation using CEPS data:

$$
\begin{aligned}
Y_{icgj} = \beta_0 + \beta_1 FS_{icgj} + \beta_2 FT_{cgj} + \beta_3 (FS_{icgj} * FT_{cgj}) + \gamma_0 LPA_{icgj} + \gamma_1 (LPA_{icgj} * FS_{icgj}) + \\
\gamma_2 (LPA_{icgj} * FT_{cgj}) + \gamma_3 [LPA_{icgj} * (FS_{icgj} * FT_{cgj})] + \beta_4 SC_{icgj} + \beta_5 TC_{cgj} + \eta_{gj} + \varepsilon_{icgj}
\end{aligned}
\tag{3}
$$

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[12]. 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

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

(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 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.

Our model generates clear predictions for three parameters. The first is γ_3, which we interpret as a quasi-experimental estimate of the effect of being assigned a female math teacher on low-perceived ability girls relative to the effect for low perceived ability boys, i.e., the effect of teacher-student gender match on the "gender gap" for low perceived ability girls (Dee, 2007; Muralidharan and Sundararaman, 2011; Lim and Meer, Forthcoming). Prediction 1 is that this coefficient 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 β_3, the effect on the gender gap for non-low perceived ability children. Prediction 2 of our model is that β_3 should be substantially smaller in magnitude than γ_3. 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 3, and we will explicitly address a few of these in the results. First, $\gamma_2 + \gamma_3$ yields the total effect on low perceived ability girls of being assigned a female 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., the comparison group is now all boys, not only low perceived ability boys).

If our assumption of orthogonality is satisfied, estimating equation 3 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[13]. We conduct two regressions - one without any fixed effects, and a second with the grade-by-school

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

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[14]. 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). Table A.3 reports the estimation results from an empirical test similar to that in Table 1, only conducting 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, 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 the teacher's gender.

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

[14] 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.

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 wherein the administrators are seeking to maximize the performance of the best students. In short, bias from this misreporting would push our coefficient estimates in the opposite direction of our framework's main predictions.

The CEPS asks students how difficult they found learning math in the sixth (and final) grade of primary school, and we use this question to proxy for students' perception of their ability[15]. 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[16]. We link this classification decision to our model by the following assumption: $E(r^i - \underline{r} \mid A_1^i = L) < E(r^i - \underline{r} \mid A_1^i = H)$. In prose: 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. This assumption implies that students classified as low perceived ability are thus closer to the margin of deciding whether or not to invest in human capital. In Table A.4 we show characteristics of students, by gender, for both of the perceived ability groups. 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%).

It is important to note that this 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 recall of perceived ability, 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

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

[16]Later in this section we discuss how alternative methods of classifying who is low perceived ability yield similar empirical results.

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.

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.2, 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 test scores between genders in either group.

To show that our analysis is not sensitive to our definition of low perceived ability, we present results from two alternative specifications in the appendix. In the first, we redefine low perceived ability to include those who report sixth grade math as "somewhat difficult" in addition to those who reported it as "very difficult." In the second specification, we replace the interaction terms for low perceived ability with an interaction for those whose math test score is below the median value of their teacher-gender pairing group (e.g., those boys paired with a male teacher whose test score is below the median score for that group). These results are somewhat smaller in magnitude than the results from the core specification, but they have the same signs and largely retain their statistical significance.

5 The effects of being assigned a female math teacher

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 student beliefs and aspirations. We then look at how this match affects investment in human capital, followed by analysis of its impact on performance on math tests.

5.1 Beliefs and aspirations

In this subsection, we conduct a test of the model's prediction that being assigned a female math teacher should positively affect beliefs and aspirations for low perceived ability girls. We investigate the impact of teacher gender on three belief variables: perceived difficulty of current math class, the careers to which students aspire, and anti-girl stereotypes. Our specification follows equation 3, using grade-by-school fixed effects and the full battery of controls for students and teachers. We also control for students' math test scores, allowing us to compare changes in beliefs while controlling for performance.

For the analysis of perceived difficulty, we use the present-time analogue to the baseline perceived ability question, students' response to the prompt "how difficult do you find your current math class to be?"[17] 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[18]. Our theory predicts the clearest break between jobs which are more often regarded as feminine/associated with women, and everything else. 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 three and equal to zero otherwise[19]. For stereotypes, we estimate the effect of teacher-student gender match on whether the student

[17]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.

[18]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.

[19]We find no effects on girls' aspirations to jobs in STEM fields. One potential explanation for this is that because of the selectivity of work in STEM, low ability children in our study are not on the margin of aspiring to work in those fields.

agrees with a statement that boys are better than girls at learning math.

We present our results in Table 2. In column 1, our results suggest that being taught by a female math teacher reduces low perceived ability girls' probability of perceiving math as "very difficult" by 20 percentage points (γ_3). While the estimated effect for non-low perceived ability girls assigned to a female teacher is the same sign as for the low perceived ability girls, it is an order of magnitude smaller and not statistically significant (β_3). Being assigned a female math teacher is also associated with an increase in low perceived ability boys' perceived difficulty of math (γ_2). These results accord with predictions 1-3 from the model.

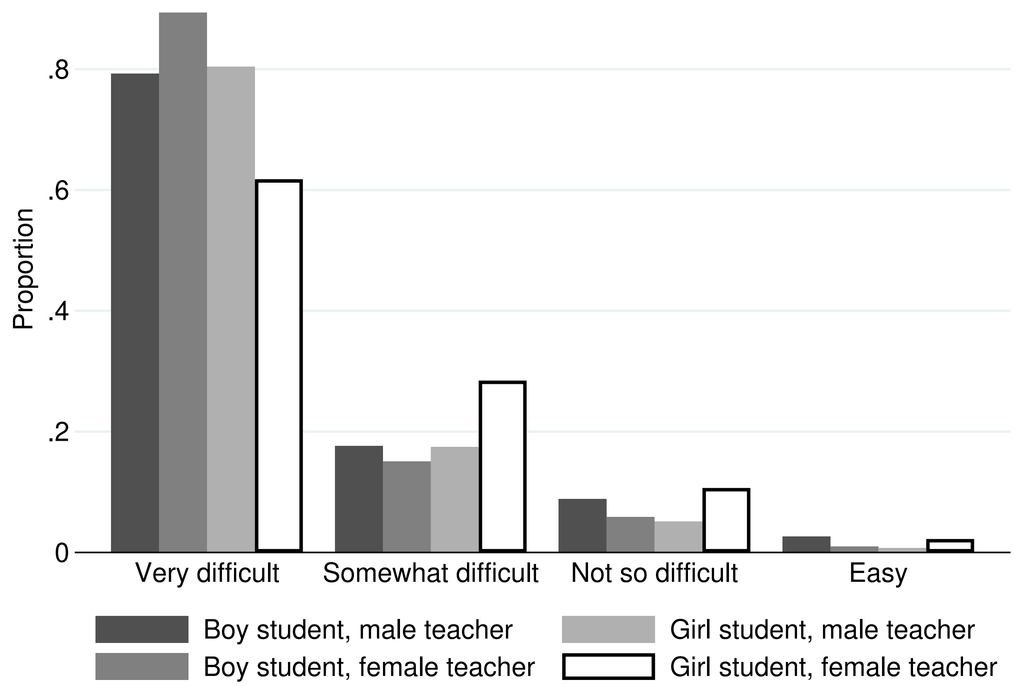
We also present results for low perceived ability students visually in Figure 3. In this figure, we plot the distribution of perceived difficulty of the current math class for each possible teacher-student gender pairing, restricting the sample to low perceived ability children. This shows the same pattern as the coefficients - 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 boys assigned to female math teachers are at least 10 percentage points more likely to find math very difficult than any other group. Table A.5 shows these results for the alternative specification of low perceived ability, where we observe a 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.3 gives the below-median analogue to Figure 3 and displays a similar pattern. We observe similar graphical results for the alternative specification of low perceived ability.

Table 2: Effects on beliefs and aspirations

	(1) Current math class perceived as very difficult	(2) Aspires to jobs in art and design	(3) Holds anti-girl stereotypes
Girl x female teacher x low perceived ability	-0.205***	-0.110**	-0.038
	(0.057)	(0.056)	(0.070)
Female teacher x low perceived ability	0.100**	-0.031	0.079
	(0.046)	(0.034)	(0.058)
Girl x female teacher	-0.037**	0.008	-0.047
	(0.017)	(0.019)	(0.035)
Girl x low perceived ability	0.046	0.105***	0.351***
	(0.042)	(0.034)	(0.053)
Girl	0.051***	0.184***	-0.130***
	(0.013)	(0.017)	(0.030)
Female teacher	0.010	0.004	0.045
	(0.017)	(0.018)	(0.032)
Low perceived ability	0.451***	-0.003	-0.196***
	(0.037)	(0.026)	(0.044)
Mean for non-LPA boys	0.122	0.104	0.599
Number of observations	8,276	8,213	8,117

Notes: The regression specification used is given in equation 3, adding a control for the student's math test scores. Point estimates and their precision are largely unchanged by removing this final control. All 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 3: Low perceived ability students' current perception of the difficulty of math, by gender of student and math teacher



Notes: This figure plots the response of low perceived ability students to the prompt: "how difficult do you find your current mathematics course to be?" This 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.

In column 2 of Table 2, we present estimates of the effect of being assigned a female math teacher on students' career aspirations. 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 positive and significant coefficients on the "girl" and "girl x low perceived ability" variables corroborate our choice of variable coding - girls, and particularly low perceived ability girls, are more likely to aspire to these jobs, independent of the gender of their teacher. 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.

In column 3, our estimate shows a small, statistically insignificant effect on girls' stereotypical beliefs. The total effect on the gender gap, $\beta_3 + \gamma_3$, however, is 8.5 percentage points, or a 13.4% decrease in the proportion of low perceived ability girls in our sample who hold these beliefs. In column 3 of Table A.6, we see a coefficient on girl x female teacher x below median of 8.0 percentage points, significant at the 10% level, and a total effect of 11.7 percentage points, or 25% of the baseline proportion. These results lead to two conjectures: one, that it may be harder to change global beliefs (stereotypes) than local beliefs (perceptions of own ability, as proxied by perceived difficulty); and two, that those with somewhat higher perceived ability may be more prone to updating their global beliefs when presented with a positive role model. In the context of our model, the mapping from a child's prior to the size of her update, as in Figure A.1, may be less right-skewed for stereotypes than for perceived ability.

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 girls matched with female math teachers. 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 0.9, 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 data, 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. We explore this gap in test scores further in the next section.

5.3 Academic performance

In this subsection we examine the effect of teacher-student gender match on students' performance in mathematics, presenting a series of analyses to further quantify the distributional differences in

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	0.091*	3.057***	0.392	0.000
x low perceived ability	(0.052)	(1.253)	(1.595)	(0.024)
Female teacher	-0.082**	-1.548	0.687	-0.014
x low perceived ability	(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 3. 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.

children's scores on midterm math examinations shown in Figure 2. We present our main results in Table 4. The first column shows the estimates generated from a version of equation 3 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 3. This is also the specification used to generate the estimates shown in Tables 2, 3, and all subsequent tables.

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 effect of teacher-student gender match on girls. In column 2, we see that the benefits estimated in column 1 appear to accrue entirely to low perceived ability girls. We find that having 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 ($\beta_3 = 0.0068$, $\sigma = 0.541$). 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. If we remove the LPA term and its interactions (i.e., all of the variables with γ coefficients), our estimate of β_3, the coefficient on teacher-student gender match, is 0.093 SD ($\sigma = 0.063$). This is well within the range of estimates generated in previous work (e.g., Dee, 2007; Muralidharan and Sundararaman, 2011; Lim and Meer, Forthcoming). 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).

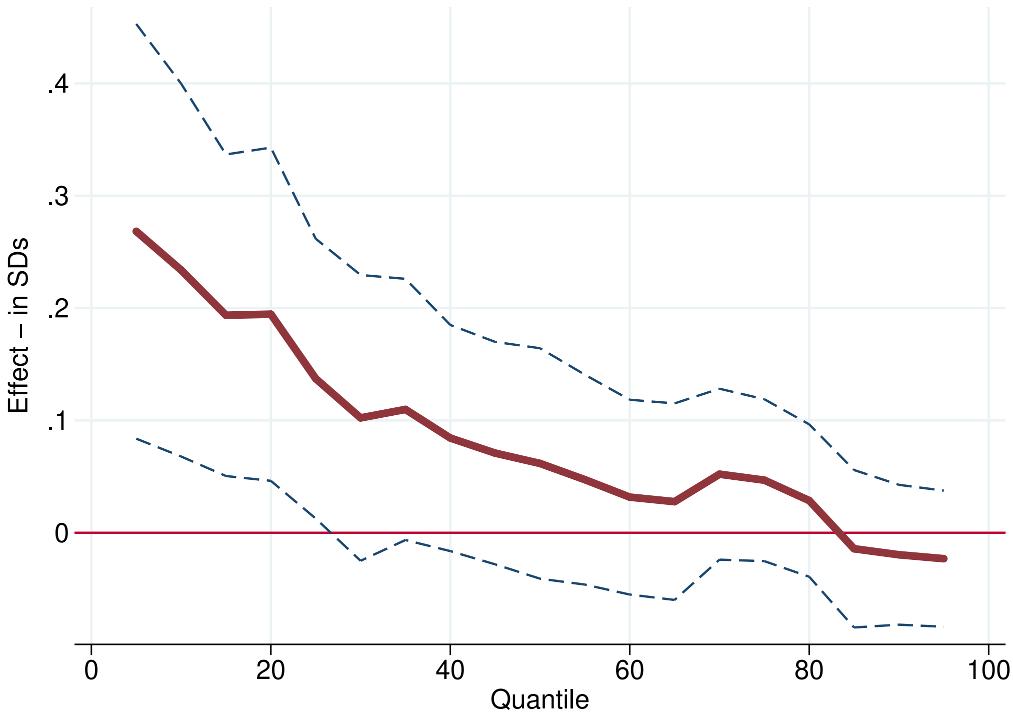
We present quantile regression results in Figure 4. To generate this figure, we estimate equation 3 without any of the independent variables related to low perceived ability and recover coefficient estimates of β_3 and the corresponding confidence interval at every fifth centile between the fifth and

Table 4: Effects on math test score

	(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 3. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

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 3 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.

95th. Consistent with our visual inspection of Figure 2, the results of the distributional difference test we conducted in Section 2, and the pattern that we see in Table 4, 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.

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 on 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 Mechanisms and discussion

In this section, we first conduct a series of analyses to test for evidence of two potential drivers of the patterns we observe in the previous section - the first, described in our conceptual framework, is that same-gendered math teachers serve as role models who counter the negative effects of stereotypes on student beliefs; the second is that there is some other characteristic of female teachers or their conduct which drives these results. 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

6.1 Mechanisms

In this subsection we first show that further exposure to female role models in mathematics has additional positive effects on the outcomes we study. We then perform a series of analyses looking for evidence of alternative explanations for the patterns we observe. The intiution 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 lavish more attention on 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 female teachers exerting more effort than male teachers drives the effects; and four, that our findings are driven by differences in teaching methods between female and male teachers.

We first show that additional exposure to the role model increases the impact of teacher-student gender match on the outcomes we measure. We exploit the fact that for some students, their math teacher is also their homeroom teacher, and these students spend additional time each day in the presence of that teacher. This provides variation in the amount of students' exposure to the role model. Our framework predicts that this additional exposure should generate additional positive effects, and we test this by estimating the additional effect of a student being assigned a female math teacher who is also the student's homeroom teacher on the same dependent variables: beliefs, stereotypes, enrollment in math tutoring, and performance on the midterm math exam. We present these results in Table A.7; while the results are imprecise, the coefficients are large and, for all but the tutoring variable, in the predicted direction.

Next, we present results from a series of tests that consider alternative explanations for the patterns we observe in Section 5. We test four possibilities: one, that female math teachers may choose to engage more with low perceived ability girls than do male math teachers; two, that female math teachers may simply be better at teaching those of low perceived ability than male teachers; three, that it is differential teacher effort instead which drives the results we observe; and four, that female teachers teach differently than do male teachers, and this difference drives our estimates.

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 3 using these two measures as outcome variables[20]. 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, reverting to the original specification in equation 3 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[21]. 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 our estimate of γ_3.

Next, we investigate the possibility that teacher effort drives our results. The CEPS collects

[20]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.

[21]Results in tabular form are available from the authors but not included in this manuscript.

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 x low perceived ability	0.035 (0.080)	-0.083 (0.069)
Female teacher x low perceived ability	-0.084 (0.055)	-0.022 (0.043)
Girl x female teacher	0.008 (0.025)	0.030 (0.024)
Girl x low perceived ability	0.012 (0.053)	0.022 (0.048)
Girl	-0.030 (0.022)	-0.055*** (0.020)
Female teacher	0.057 (0.035)	0.024 (0.036)
Low perceived ability	-0.077* (0.043)	-0.097*** (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 3, 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) Believes boys are better than girls at learning math	(4) Midterm math test score
Girl x award-winning teacher	0.040	0.105	-0.079	-0.072
x low perceived ability	(0.068)	(0.068)	(0.068)	(0.168)
Award-winning teacher	-0.037	-0.018	0.027	0.157
x low perceived ability	(0.055)	(0.035)	(0.054)	(0.134)
Girl x award-winning teacher	-0.012	-0.064***	0.037	0.049
	(0.015)	(0.019)	(0.034)	(0.052)
Girl	-0.070	0.006	0.371***	0.229*
x low perceived ability	(0.049)	(0.042)	(0.052)	(0.133)
Girl	0.033***	0.219***	-0.176***	0.108***
	(0.010)	(0.015)	(0.024)	(0.037)
Award-winning teacher	0.005	0.055***	-0.010	-0.081
	(0.021)	(0.019)	(0.040)	(0.092)
Low perceived ability	0.519***	-0.013	-0.167***	-0.956***
	(0.039)	(0.024)	(0.043)	(0.099)
Mean for non-LPA boys	0.122	0.104	0.599	7.024
Number of observations	8,276	8,213	8,117	8,294

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

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 the effort variable. 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. In Tables A.8, A.9, and A.10, we show results for the same dependent variables: perceived difficulty of math, whether the child holds anti-girl stereotypes, enrollment in math tutoring, and the math midterm exam score. For none of these analyses do we observe a significant effect of teacher effort on outcomes for low perceived ability girls[22].

Finally, we look at the impact of teachers' use of different methods of teaching on low perceived ability girls. 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) rather than the role model effect. The CEPS records teachers' response to the following question - "how often do you use [teaching method]: never, sometimes, often, or always?" - 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.11 shows estimates of the effect of teachers' use of these methods on perceived

[22]We cannot entirely exclude an alternative explanation for this pattern: teachers who expend less effort in terms of out of class hours may merely be more productive with their time.

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.4, and note that the large difference seen in the math scores of low perceived ability girls assigned to female teachers is not apparent; overall, we see some evidence of (substantially smaller) effects of teacher-student gender match on beliefs and test scores. Whether because of smalll 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 - or both, 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 children in 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. We model this process and take our model's predictions to an empirical context where strong negative stereotypes regarding girls' math ability vis-a-vis boys' prevail despite girls' superior performance in math. Our empirical results bear out the model's predictions. We find that low perceived ability students benefit from being assigned a same-gendered teacher, and these benefits are largest for low perceived ability students who also face stereotypes, i.e., girls. Same-gendered 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, anti-girl stereotypes may persist and harm the most vulnerable. We show that dosing these students with shared-identity role models can dampen and, for some outcomes, reverse the negative effects of stereotypes. Second, it adds to a growing body of research (e.g., Bernard et al., 2014; Lusher et al., 2015; Lybbert and Wydick, 2016b; Genicot and Ray, 2017) in economics studying how information affects aspirations, educational decisions and outcomes. 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. More broadly, our study suggests that role models and the information they provide are likely an important input into the production of human capital, particularly among girls and other groups who for various historical or socioeconomic reasons may lack for a credible example of successful investment in certain types of human capital.

References

Akerlof, G.A. and R.E. Kranton, "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature*, 2002, *40* (4), 1167–1201.

Akerlof, George A. and Rachel E. Kranton, "Economics and identity," *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

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.

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.

Bénabou, Roland and Jean Tirole, "Identity, morals, and taboos: Beliefs as assets," *Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

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.

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.

Fortin, Nicole M, Philip Oreopoulos, and Shelley Phipps, “Leaving boys behind gender disparities in high academic achievement,” *Journal of Human Resources*, 2015, *50* (3), 549–579.

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

Gershenson, Seth, Stephen B Holt, and Nicholas W Papageorge, “Who believes in me? The effect of student–teacher demographic match on teacher expectations,” *Economics of Education Review*, 2016, *52*, 209–224.

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.

Jensen, Robert, "The (perceived) returns to education and the demand for schooling," *Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

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.

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.

Lai, Fang, Elisabeth Sadoulet, and Alain De Janvry, "The contributions of school quality and teacher qualifications to student performance evidence from a natural experiment in Beijing middle schools," *Journal of Human Resources*, 2011, *46* (1), 123–153.

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

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*, 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," *University of California, Davis Working Paper*, 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 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.

Ross, Phillip H, "Aspirations and human capital investment: Evidence from Indian adolescents," *Working Paper*, 2016.

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.

Sherman, David K, Kimberly A Hartson, Kevin R Binning, Valerie Purdie-Vaughns, Julio Garcia, Suzanne Taborsky-Barba, Sarah Tomassetti, A David Nussbaum, and Geoffrey L Cohen, "Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat.," *Journal of Personality and Social Psychology*, 2013, *104* (4), 591–618.

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, Claude M, Steven J Spencer, and Joshua Aronson, "Contending with group image: The psychology of stereotype and social identity threat," *Advances in Experimental Social Psychology*, 2002, *34*, 379–440.

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.

Zhang, Hongliang, "Identification of treatment effects under imperfect matching with an application to Chinese elite schools," *Journal of Public Economics*, 2016, *142*.

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
Grade-by-school fixed effects		X
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]

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) Believes boys are better than girls at learning math	(4) Midterm math test score
Girl x female teacher x low perceived ability (alternate definition)	-0.113***	-0.053	-0.040	0.274***
	(0.040)	(0.036)	(0.047)	(0.109)
Female teacher x LPA alternate definition	0.027	0.012	0.009	-0.110
	(0.032)	(0.018)	(0.035)	(0.086)
Girl x female teacher	-0.025	0.014	-0.030	-0.039
	(0.016)	(0.022)	(0.035)	(0.061)
Girl x LPA alternate definition	0.078***	0.076***	0.304***	-0.065
	(0.031)	(0.026)	(0.036)	(0.077)
Girl	0.021	0.169***	-0.208***	0.212***
	(0.013)	(0.019)	(0.030)	(0.053)
Female teacher	0.017	0.000	0.048	0.190***
	(0.018)	(0.018)	(0.031)	(0.067)
Low perceived ability (alternate definition)	0.180***	-0.015	-0.102***	-0.647***
	(0.025)	(0.014)	(0.027)	(0.064)
Mean for non-LPA boys	0.094	0.100	0.615	7.137
Number of observations	8,276	8,213	8,117	8,294

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *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) Believes boys are better than girls at learning math	(4) Midterm math test score
Girl x female teacher x below median	-0.078***	0.009	-0.080*	0.111*
	(0.027)	(0.034)	(0.046)	(0.057)
Female teacher x below median	0.007	-0.004	0.045	0.087***
	(0.024)	(0.019)	(0.034)	(0.033)
Girl x female teacher	-0.039**	-0.013	-0.028	0.012
	(0.017)	(0.022)	(0.041)	(0.029)
Girl x below median	0.068***	0.026	0.287***	0.056
	(0.022)	(0.026)	(0.034)	(0.046)
Girl	0.058***	0.188***	-0.224***	0.047**
	(0.014)	(0.019)	(0.034)	(0.024)
Female teacher	0.026	0.006	0.036	0.061**
	(0.020)	(0.018)	(0.039)	(0.030)
Below median	-0.031	-0.002	-0.112***	-1.653***
	(0.021)	(0.023)	(0.031)	(0.029)
Mean for above median boys	0.069	0.085	0.638	7.757
Number of observations	8,300	8,251	8,151	8,345

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.7: Effect of having math teacher as homeroom teacher

	(1) Perceived difficulty of current math class	(2) Believes boys are better than girls at learning math	(3) Enrolled in after-school math tutoring	(4) Midterm math test score
HRMT x female teacher x girl x LPA	-0.113	-0.284**	-0.187	0.145
	(0.128)	(0.131)	(0.121)	(0.356)
HRMT x female x girl	0.027	0.053	-0.036	-0.017
	(0.034)	(0.078)	(0.047)	(0.112)
HRMT x LPA x girl	0.073	0.157	0.042	0.095
	(0.082)	(0.109)	(0.081)	(0.229)
Homeroom teacher = math teacher (HRMT) x LPA)	-0.033	-0.100	-0.068	-0.189
	(0.083)	(0.108)	(0.050)	(0.181)
HRMT x LPA	0.010	-0.003	0.169**	-0.186
	(0.102)	(0.132)	(0.078)	(0.258)
HRMT x female teacher	-0.084**	0.068	-0.048	0.347**
	(0.042)	(0.072)	(0.056)	(0.156)
HRMT x girl	-0.022	-0.087	0.014	-0.035
	(0.026)	(0.060)	(0.036)	(0.096)
Girl	0.057***	-0.100***	0.018	0.137***
	(0.017)	(0.034)	(0.019)	(0.055)
HRMT	0.008	0.022	0.041	-0.088
	(0.031)	(0.054)	(0.037)	(0.112)
Female math teacher	0.029	0.029	0.008	0.079
	(0.023)	(0.033)	(0.032)	(0.079)
Low perceived ability	0.464***	-0.161***	0.065*	-0.738***
	(0.038)	(0.055)	(0.034)	(0.090)
Mean for non-LPA boys	0.122	0.599	0.210	7.024
Number of observations	8,276	8,117	8,257	8,294

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.8: Teacher effort 1

	(1) Perceived difficulty of current math class	(2) Believes boys are better than girls at learning math	(3) Enrolled in after-school math tutoring	(4) Midterm math test score
Hours prep: hours in class x low perceived ability x girl	0.037	-0.068*	0.021	-0.062
	(0.056)	(0.039)	(0.033)	(0.123)
Hours prep: hours in class x low perceived ability	-0.018	0.032	-0.006	0.152**
	(0.032)	(0.034)	(0.024)	(0.076)
Hours prep: hours in class x girl	-0.007	0.003	-0.011	0.051
	(0.007)	(0.015)	(0.012)	(0.032)
Girl x low perceived ability	-0.093	0.406***	-0.030	0.283
	(0.073)	(0.058)	(0.055)	(0.177)
Girl	0.036***	-0.161***	0.051***	0.075
	(0.011)	(0.026)	(0.018)	(0.048)
Hours prep: hours in class	0.019*	0.001	-0.033***	-0.097*
	(0.011)	(0.020)	(0.010)	(0.058)
Low perceived ability	0.520***	-0.192***	0.003	-1.055***
	(0.046)	(0.042)	(0.031)	(0.108)
Mean for non-LPA boys	0.122	0.599	0.210	7.024
Number of observations	8,212	8,055	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.9: Teacher effort 2

	(1) Perceived difficulty of current math class	(2) Believes boys are better than girls at learning math	(3) Enrolled in after-school math tutoring	(4) Midterm math test score
Hours grading: hours in class x low perceived ability x girl	0.013	-0.024	0.022	-0.099
	(0.042)	(0.041)	(0.050)	(0.130)
Hours grading: hours in class x low perceived ability	-0.053	0.015	0.000	0.149*
	(0.036)	(0.034)	(0.030)	(0.087)
Hours grading: hours in class x girl	-0.011	-0.011	0.006	0.051*
	(0.007)	(0.019)	(0.012)	(0.030)
Girl x low perceived ability	-0.071	0.357***	-0.031	0.319*
	(0.064)	(0.063)	(0.055)	(0.186)
Girl	0.041***	-0.145***	0.034**	0.075*
	(0.011)	(0.029)	(0.017)	(0.045)
Hours grading: hours in class	0.014	0.059***	-0.004	0.075
	(0.013)	(0.025)	(0.016)	(0.057)
Low perceived ability	0.559***	-0.171***	-0.004	-1.050***
	(0.051)	(0.045)	(0.034)	(0.122)
Mean for non-LPA boys	0.122	0.599	0.210	7.024
Number of observations	8,212	8,055	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

Table A.10: Teacher effort 3

	(1) Perceived difficulty of current math class	(2) Believes boys are better than girls at learning math	(3) Enrolled in after-school math tutoring	(4) Midterm math test score
Hours prep + grading: hours in class x low perceived ability x girl	0.016	-0.034	0.016	-0.053
	(0.032)	(0.024)	(0.022)	(0.076)
Hours prep + grading: hours in class x low perceived ability	-0.024	0.017	-0.002	0.105**
	(0.022)	(0.019)	(0.016)	(0.051)
Hours prep + grading: hours in class x girl	-0.006*	-0.003	-0.001	0.033*
	(0.004)	(0.009)	(0.007)	(0.018)
Girl x low perceived ability	-0.090	0.405***	-0.040	0.334
	(0.082)	(0.068)	(0.058)	(0.212)
Girl	0.042***	-0.152***	0.043**	0.057
	(0.012)	(0.029)	(0.019)	(0.050)
Hours prep + grading: hours in class	0.012*	0.017	-0.015*	-0.020
	(0.006)	(0.013)	(0.008)	(0.035)
Low perceived ability	0.554***	-0.194***	-0.000	-1.121***
	(0.058)	(0.046)	(0.036)	(0.134)
Mean for non-LPA boys	0.122	0.599	0.210	7.024
Number of observations	8,212	8,055	8,193	8,230

Notes: The dependent variable in question is given in the column heading. Dependent variables in columns 1-3 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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

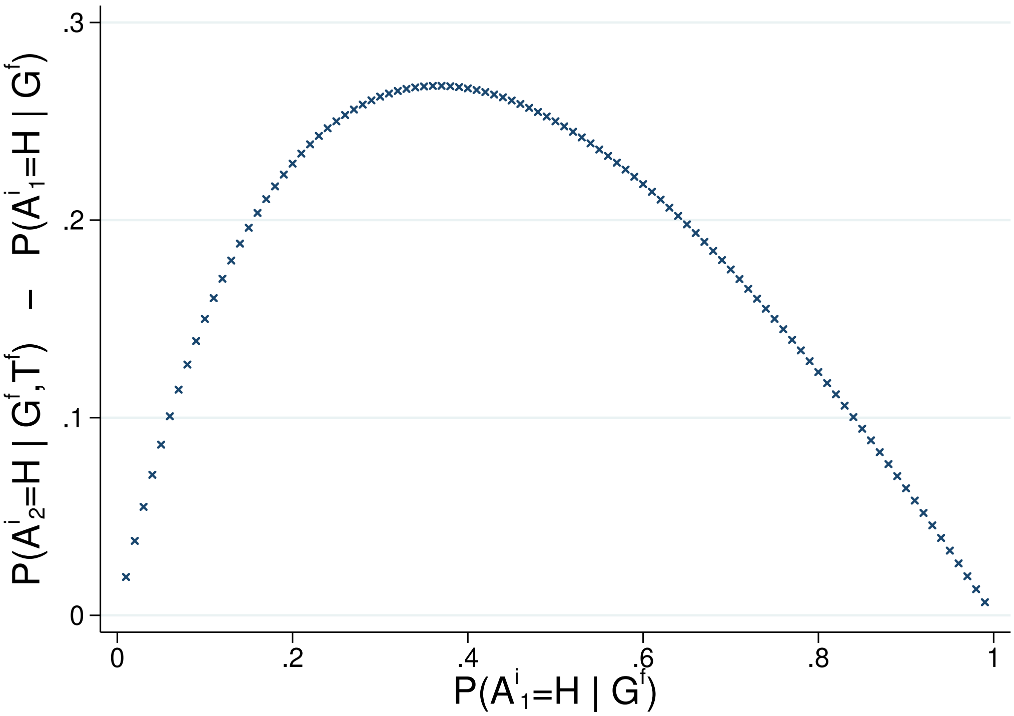
Table A.11: Teaching method

	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 3. *p < 0.1, **p < 0.05, ***p < 0.01.

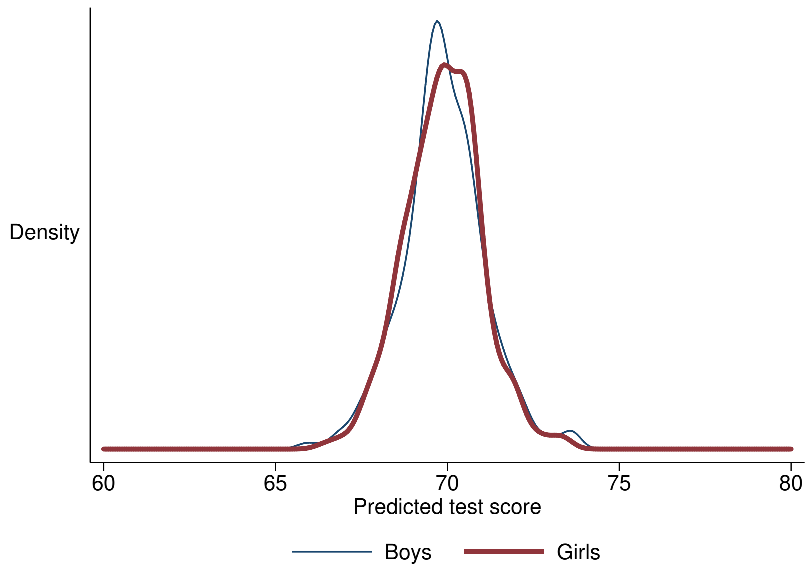
Appendix B: Appendix figures

Figure A.1: Mapping of prior to size of update

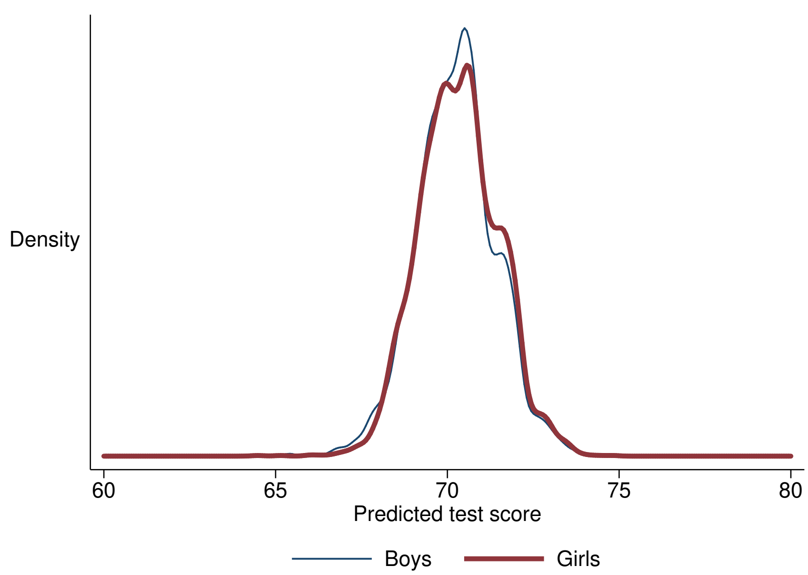


Notes: this figure shows the mapping from a girl's prior that she is of high ability, $P(A_t^i = H|G^f)$, to the update of that prior in response to encountering a female math teacher. The assumptions used to generate this figure are $P(G^f, T^f|A_t^i = H) = 0.6$ and $P(G^f, T^f|A_t^i = L) = 0.2$, but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

Figure A.2: Predicted test score distributions, by perceived ability



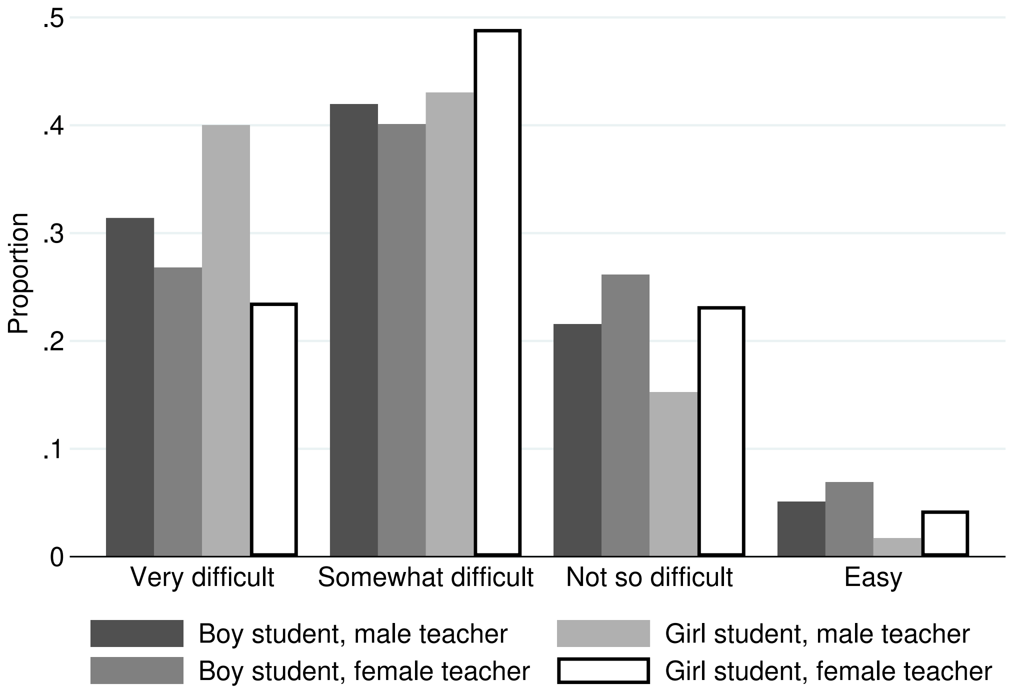
Panel A: Low perceived ability



Panel B: Not low perceived ability

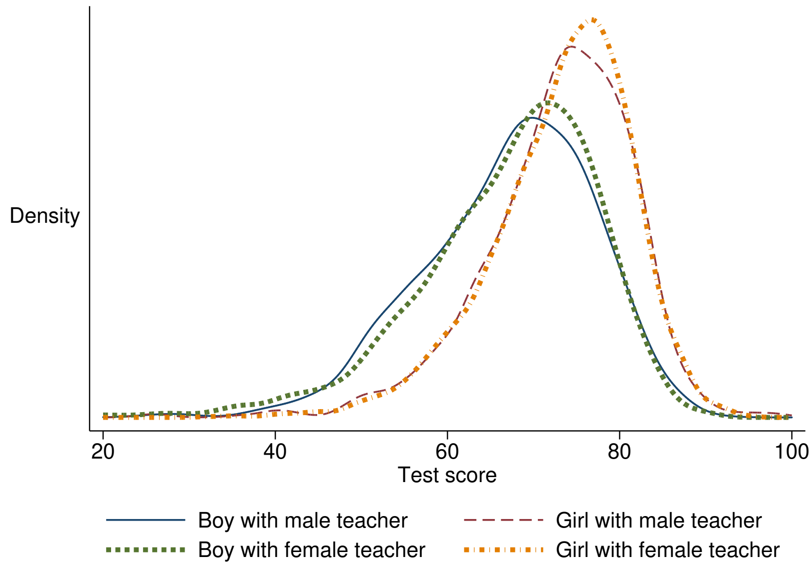
Notes: to generate these figures, we regress test scores on the vector of student-level predetermined characteristics and, using these coefficients, generate a predicted test score for each student. We then plot these using a gaussian kernel for each perceived ability-gender group.

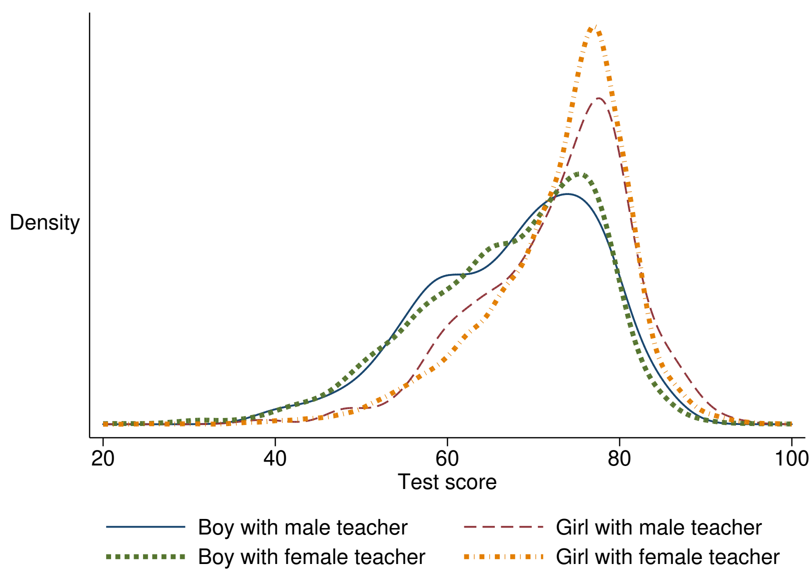
Figure A.3: Effect of teacher-student gender match on student beliefs, for those below within-group median test score



Notes: this figure shows the same analysis as reported in Figure 3, only limiting the sample instead to those below the within-group median math test score.

Figure A.4: Distribution of English and Chinese test scores by teacher-student gender pairing





Notes: this figure shows the analogue to Figure 2 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: 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.