

## **Investigating the Causal Impact of Teacher Gender Match on Student Learning**

Congli Zhang and Anwesha Guha

Quantitative Research Methods in Education

College of Education, University of Oregon

### **Author Note**

Correspondence concerning this article should be addressed to Anwesha Guha, PhD Student in Education Studies Department at the University of Oregon. Location: Lokey Bldg, 5267 University of Oregon, Eugene, OR 97403. Email: [aguha@uoregon.edu](mailto:aguha@uoregon.edu)

### **Investigating the Causal Impact of Teacher Gender Match on Student Learning**

Studies have attempted to understand the gender gap in school by examining student-teacher gender matching (Dee, 2007; Sansone, 2017). Many find a significant positive relationship of having a female-identifying teacher for girls, especially in middle school – the age that students tend to internalize gender stereotypes (Gong et al., 2018; Lim and Meer, 2017; Xu and Li, 2018). Few of these papers, however, have explored this relationship causally. In this paper, we conduct a causal analysis using the China Education Panel Survey (CEPS) by leveraging a unique policy change in China that forced teacher-student random assignment in 2006.

Our study contributes to a rigorous literature body that examines the causal impact of teacher gender matching on student academic outcomes and self-confidence. We attempt to obtain precise estimates of teacher effects in three ways. First, given random assignment enforced by national policy, we are able to use between-teacher variation to estimate internally valid teacher effects (see Ladd and Sorensen, 2017; Papay and Kraft, 2015). Second, we model teacher gender and gender matching separately to understand the role both gender effect and gender matching play on student outcomes. Third, we control for school fixed-effects to account for systematic differences across schools and a rich set of student-, homeroom-, and teacher-level covariates to improve estimation precision (see Chetty et al., 2014; Kraft, 2019). Here, the most important note is the cubic polynomial functions of four scores (in three core content subjects and cognitive test) in prior years that effectively absorb noise from individual learning ability.

We control for baseline performance differences in female students to find that even though there are gender-match differences, the gender-match effect is smaller in effect. Thus, we build on existing models to better isolate the relationship between achievement and gender equity and discuss its policy implications.

### **Background**

Teachers serve as role models to students in three key ways: as behavioral models in showing students how to perform a skill or achieve a goal, as inspirations in making certain goals desirable, and as representations of the possible (Morgenroth et al., 2015). This motivational theory of role models works alongside theories of social identity and self regulation to posit teachers as powerful influences not just on student learning, but on student confidence, performance, and experience. Role models are particularly important for girls, for whom gender disparities lead to detrimental impacts in academic performance and later life outcomes (Sansone, 2019; Terrier, 2020).

### **Investigating the Role of Gender in Learning Outcomes**

Studies have attempted to understand the gender gap in school by examining student-teacher gender matching (Dee, 2007; Sansone, 2017). However, it is important to isolate the gender impacts and match impacts associated with achievement, as discussed in the sections below.

### ***Student Gender Gap***

Socially constructed and situation-dependent identities, such as gender, have pushed groups such as girls to be subject to stereotype threat, where negative genders norms associate being a girl with lower ability to perform in subjects such as math and other STEM subjects (McIntyre et al. 2005; Spencer et al., 1999). While the impact of this stereotype threat varies in different contexts and across different ages of psychological development (Flore & Wicherts, 2015), these biases and stereotypes continue to affect schooling experiences for girls worldwide (UNESCO, 2022). For example, although girls have slowly and increasingly outperformed boys in school, particularly secondary classrooms, female students “remain much less likely to major in quantitative, technical, and science-related fields” – furthering segregation by occupation and inequality of income by gender in the long term (Bettinger & Long, 2005, p. 152).

Additionally, gender differences impact boys in different ways. For example, research suggests that boys exhibit higher levels of externalizing behavior problems than girls, even through middle school (Castelao & Kroner-Herwig, 2014) – behavior that is strongly associated with lower school achievement (Kulkarni et al., 2020). In addition, gender stereotype threat similarly affects boys, discouraging confidence and achievement when negative stereotypes are communicated (Hartley & Sutton, 2003).

### ***Teacher Gender Impact***

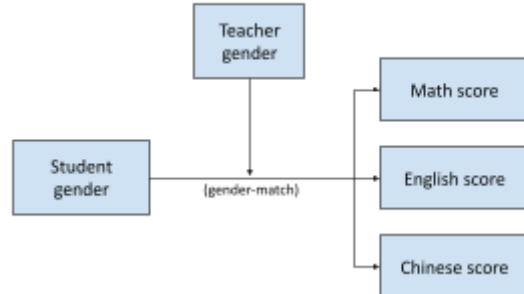
Given differences in experiences for students, teacher gender might matter for student cognitive and noncognitive outcomes, particularly in the way teachers affect student interest and self-efficacy (Sansone, 2017). Using teacher value-added measure as proxy of teacher quality, Aaronson et al. (2007) found that female teachers are associated with about 0.07 grade equivalents higher than male teachers and this difference derives from female students, especially from female students with similar demographics. Antecol et al. (2015) finds that having a female teacher lowers math scores for female students, though that may be due to students assigned female teachers with limited math background. These effects indicate while teacher gender matters, that characteristic becomes insignificant once behaviors and attitudes about ability and gender are taken into account (Sansone, 2017). So, we follow the theoretical framework set by Paredes (2014) in that teacher gender may affect students in a variety of ways: by acting as role models, reinforcing stereotype threats, and through teacher biases that are correlated with gender (Paredes, 2014; Lavy, 2008).

### ***Gender Match Effects***

Gender match effects literature is mixed. While studies find a significant positive relationship of having a female-identifying teacher for girls, especially in middle school (Gong et al., 2018; Lim and Meer, 2017; Xu and Li, 2018), other research has found different outcomes. For example, Cho (2012) used a student fixed-effects model to estimate the effects of teacher-student gender matching using

math and science data from 15 OECD countries and found that, in most countries, neither boys nor girls benefit from gender match.

Figure 1. Gender-match effects framework



We are interested in the mechanisms and impacts of gender match effects specifically in China, particularly for its student- and teacher- random assignment policy.

Research has found significant differences in perceived ability to succeed by gender in subjects like math (Tsui, 2007), with boys exhibiting significantly higher growth mindsets and self efficacy than girls (Su et al., 2021). However, there are gaps in understanding how achievement and ability are influenced by teachers – the role models in the classroom. Both Xu and Li (2018) and Gong et al. (2018) explore the causal impact of teacher gender match on student learning by leveraging the unique policy change in China that forced teacher-student random assignment in 2006. Using either school- and school-by-grade level fixed effects models, authors find significant effects of having a female subject teacher (English, math and Chinese) for girls but not for boys.

With these differences in results in research, Eble & Hu (2020) attempts to disentangle the active and passive mechanisms at play in the student-teacher gender match by building on identity theory, which Dee (2005) notably explores. Their results support the role-modeling theory, confirming that “the intersection of a child’s beliefs about themselves and societal beliefs about ability by gender” importantly predicts incidence and size of teacher-student gender match effects (p. 16). Thus, the student gender gap, teacher gender impact, and gender-match effects are important to disentangle in both research and policy.

### Policy Context

This study is possible due to a unique policy change in China. In 2006, in an effort to prioritize education equality, the Compulsory Education Law disallowed the sorting of students to teachers based

on student academic performance at elementary (grades 1-6) and middle school (grades 7-9) levels. In order to comply, many local education departments create either random or stratified homerooms of students in schools, standardizing and the number of teacher groups, then randomly assigning teacher groups to student homerooms. This random assignment is crucial to our identification strategies and will be discussed in further detail in the Methods section.

Unlike the US, China uses a homeroom-based school system (similar to France, Germany, India, Japan, Netherlands, Russia, and South Korea), where students are grouped into homerooms, put on a shared homeroom schedule, and assigned a group of subject teachers who rotate to the homeroom to teach. Throughout all years in which they attend the same school, students typically remain grouped with the same homeroom cohort and their core content teachers are encouraged to follow the homeroom rising up grades. This school system, compared to the one in the US where student tracking and selecting to teachers are common, makes the random teacher-student assignment more feasible and manageable in terms of both school practice and policy regulation – minimizing differential peer effects. This allows us to methodologically isolate the teacher-student relationship from other observed and unobserved confounding factors and unbiasedly answer the research questions for this paper.

Furthermore, in this paper, we focus on middle school years for 1) its importance as a child developmental period, and 2) data availability. In China, middle schools prepare students for a high-stakes high school admission exam administered by a city-level education agency. This intends to place students in different tracks: those who pass certain cutoffs (approximately half nationwide) will be admitted to general high schools that are traditional pathways to academic-focused higher education, while others will be assigned to vocational or alternate schools that prepare them for the job market. Given the high stakes of the high-school entrance exams, almost all middle schools in the nation test students at least twice (mid-semester and end-of-semester) per semester – four times a year – on core subjects such as Chinese (language arts), English (national mandated second language at school), and

math (comprehensive mathematics). These test scores and students' knowledge about their performance – knowledge developed through constant monitoring and reflecting on these scores – can together be rich outcome measures of student academic achievement. In addition, we also have access to students' self-reported reflections on their learning environment which allows us to measure their attitudes towards their learning.

### **Research Questions**

Using the CEPS data, this paper addresses the following questions:

- Whether and to what extent does being taught by a female teacher impact student academic outcomes in different subject areas including Chinese, English, and math in Chinese middle schools?
- Whether and to what extent does teacher-student gender match impact student academic outcomes in different subject areas including Chinese, English, and math in Chinese middle schools?

### **Method**

#### **Data and Measures**

##### ***China Education Panel Survey***

We drew our analytic sample from China Education Panel Survey (CEPS), China's first nationally representative, longitudinal survey of middle-school students. Starting in the 2013-2014 school year, the CEPS team implemented a stratified, multi-stage sampling scheme to randomly select 112 middle schools (104 are public) from across the country. Administrators from each randomly selected school were surveyed. Within each middle school, the sampling scheme then also selected two 7th grade and two 9th grade homerooms to survey. Within each homeroom, all students, parents, teacher-advisors, and content teachers in three core subjects (Chinese, English, and math) were surveyed. In the

2014-2015 school year, most of the initial 7th grade cohort were successfully followed up in 9th grade ( $N = 9,449$ , 91.93%). These two waves of students were the primary focus of our analysis.

The two-wave CEPS data contained not only longitudinal information on a rich set of student-, family-, teacher-, and school-level variables but also whether the school randomly assigns teachers to students. Specifically, in a baseline school survey, administrators were asked whether the school had randomly assigned teachers to students upon students' entry to middle school (before 7th grade began). This variable, coupled with the national random assignment stimulated by the 2006 Compulsory Education Law, has been leveraged by researchers to overcome selection bias in estimating student outcomes and add causal evidence to a variety of educational research areas such as teacher-student identity match (Eble & Hu, 2020; Gong et al., 2018; Xu & Li, 2018) and after-school tutoring (Sun et al., 2020).

### ***Data Characteristics***

**Sample Restriction.** Because the random teacher-student assignment is central to our identification strategy, there was a concern that even with the national policy enacted, under-resourced schools in less-reviewed areas may still sort students to teachers. To deal with this issue, we built upon prior studies (Eble & Hu, 2020; Gong et al., 2018; Xu & Li, 2018; Garrett & Steinberg, 2015) and implemented more careful restriction criteria to obtain our final analytic sample. We first limited sample schools to 85 schools that were public schools and self-reported to have randomly assigned teachers to students in the baseline survey. We then moved to address student sorting between Wave 1 and Wave 2 because more than 80% of homerooms had at least some change in their membership during the year. Most of these changes were driven by students moving from one school to another.

For identification purposes, we were relatively unconcerned about across-school sorting because in all the models we fit, we controlled for school fixed effects to absorb any time-invariant factors driving students to sort in or out of school. In contrast, we were concerned about within-school sorting, which



would introduce considerable bias into the estimates of teacher effects. To reduce this threat, we further excluded schools that had at least one student change homeroom ID (but remain in the same school) between two waves. We were left with 63 schools, which we used in our primary analyses. As a check of our findings, we returned to the 85-school sample (see Xu & Li, 2018) to conduct robustness checks and see if our findings hold.

**Missing Values.** We obtained three separate samples by matching students with their Chinese, English, and math teachers. We dropped all observations that had any missing predictor or outcome variables (1.7% missing or less) and replaced missing values on other variables with leave-one-out mean within homeroom (for student variables) or school (for teacher variables). This left us with samples of 4,945 Chinese, 4,962 English, and 5,016 math students.

### **Data Analysis Plan**

#### ***Validity of Random Assignment***

**Quality of Measures.** For each subject, student academic outcomes were measured using two variables: the student's score on the school-administered mid-fall semester exam and self-reported confidence level, measured on a four-point Likert scale to report their perception of the subject (0 = very difficult, 4 = not difficult at all). Both were standardized to have a mean of zero and standard deviation (SD) of one within each school. Teacher background and demographic information were collected from the teacher survey. Control variables were included at the student, teacher, and school level.

**Identification Strategy.** To account for the possibility of student sorting (e.g., students sorting to same gender teachers and motivated students sorting to more effective teachers), we leveraged the random teacher-student assignment that was not only enforced by the national regulation and reported by the surveyed schools, but also confirmed in the data. Specifically, following the procedure implemented by Xu and Li (2018) and Garrett and Steinberg (2015), we conducted a series of covariates balance checks by regressing two predictor variables (teacher gender and teacher-student gender

match) against all student-level covariates from baseline data while controlling for school fixed-effects and clustering standard error at the school level.

The balance check returned different results for the two predictor variables. For teacher gender, all of the coefficients on the baseline covariates were small in magnitude, and only 3 of 72 of these tests were significant at conventional levels (likely due to sample idiosyncrasy). Importantly, none of the four baseline scores appear significantly correlated with teacher gender. We conclude that the random assignment assumption required by our identification strategy was met. Provided random assignment, the variation in teacher gender was independent from any observed and unobserved factors that also impact student outcomes; therefore, the estimated change in student outcomes can be unbiasedly attributed to the treatment – being taught a year by a female teacher (compared to a male teacher).

However, this was not the case for the other predictor variable, teacher-student gender match. Rather than selection bias, this bias was generated from both baseline difference and disproportion of female teachers in the workforce in certain subject areas. Thus, gender match can be considered an endogenous and unbiased estimate.

### ***Estimation Framework***

Our strategy for estimating teacher gender effects has three major components. We began by controlling for school-level fixed effects. We identified the coefficients based on the variation in teacher characteristics within each school – effectively eliminating biases associated with student across-school sorting. Because of this strategy, in addition to the fact that our samples were drawn under a no within-school sorting criterion, we are confident that the validity of the random teacher-student assignment in our analytic samples holds. Second, we controlled for a rich set of student-, homeroom-, and teacher-level covariates from baseline data to improve estimation precision. Specifically, we included the cubic polynomial functions of baseline Chinese, English, math, and CEPS cognitive test scores in all the models we fit to account for varying functional forms of student prior learning ability and school and

family inputs (see Chetty et al., 2014; Kraft, 2019). Third, we clustered standard error at the school level to account for the within-school correlations among residuals.

### **Model Specification**

Based on our identification and estimation strategies, we recover the causal impact of teacher gender on all students by estimating a value added ordinary least squares (OLS) regression model below:

$$OUT_{ijst} = \beta_0 + \beta_1 TGENDER_{jt} + \beta_2 X_{it-1} + \beta_3 H_{jt-1} + \beta_4 T_{jt-1} + \theta_s + \varepsilon_{ijst}$$

where  $i, j, s$ , and  $t$  denote student, teacher, school, and year, respectively. Variable  $OUT_{ijst}$  is student  $i$ 's academic performance or confidence in year  $t$ ;  $TGENDER_{jt}$  is coded one if teacher  $j$  is female and zero otherwise;  $X_{it-1}$ ,  $T_{jt-1}$ , and  $H_{jt-1}$  are baseline student-, teacher-, and homeroom-level covariates;  $\theta_s$  is the school fixed effects of school  $s$ ; and  $\varepsilon_{ijst}$  is the idiosyncratic error term. The coefficient of interest is  $\beta_1$ , which is the estimated causal effect of teacher gender on student outcome.

To identify the causal impact of teacher-student gender match by student gender, we also estimate the value-added ordinary least squares (OLS) regression model below twice, for female and male students separately:

$$OUT_{ijst} = \beta_0 + \beta_1 MATCH_{jt} + \beta_2 X_{it-1} + \beta_3 H_{jt-1} + \beta_4 T_{jt-1} + \theta_s + \varepsilon_{ijst}$$

where  $MATCH_{jt}$  is coded one if teacher-student genders are the same and zero otherwise, and the coefficient of interest is  $\beta_1$ , which is the estimated causal effect of teacher-student gender match on student outcome.

We specifically do not use a random effects model because we are not interested in the school-level effects on student outcomes. We assume schools are not drastically different from each other due to observable or unobservable differences, which we ensure with our teacher-level balance checks (see Table 2 and 3).

## Results

Descriptive statistics in Table 3 show that the likelihood of a same-gender teacher match for female students is higher than for male students for Chinese, English and math, with the English sample showing the highest difference (89.77% match for female students compared to 10.85% match for male). This is due to a higher number of female teachers than male teachers (see Table 2 balance check). To isolate the effects of gender, achievement, and gender-match, we report findings isolated and then together.

### Gender Effects in Chinese and English

At baseline, female students generally perform better than their male counterparts, even when controlling for other student-level demographics and covariates (see Table 4). Naive results in Table 5 find a significant effect of teacher gender on student test scores and confidence for all main subjects, Chinese ( $SD=0.176$ ), English ( $SD=0.091$ ) and math ( $SD=-0.081$ ). However, as girls outperform boys to begin with in Chinese and English, we find that teacher-student gender match does not differentially impact boys or girls in either of these subjects, for both test score (see Table 6) and subject confidence (see Table 7).

### Gender Effects in Math

Unlike the null effects in Chinese and English, we find that teacher gender match in math leads to 0.178 standard deviation increase in test score for girls, controlling for all other covariates. There are no significantly different effects for boys (see Table 6). Additionally, there are no significant effects of gender match on confidence level, for boys, girls, or together (see Table 7).

## Discussion

This paper adds to the student-teacher gender match literature by taking advantage of a unique policy implementation and school setting in China, where teachers are randomly assigned to classrooms of students, in order to isolate teacher effects on student outcomes. We carefully draw our analytic

sample from a nationally representative dataset and provide evidence on the validity of the random assignment of teachers to students in the data. To further ensure internal validity, we control for school-level fixed-effects and cluster standard errors at school level. After controlling for baseline differences in scores, we find significant impacts of student-teacher gender match only for test scores of girls in math.

### **Implications**

While this motivational theory of teachers as role models vary across cultures (Lockwood et al., 2005) and in different situations, the power and influence teachers have on students distinguish them as key drivers in student success – especially with a demographic match. The impact of a student-teacher gender match on student learning outcomes has been increasingly salient and studied, and research has found that female mentors may be especially useful in providing counseling and emotional support (Allen & Eby, 2004; Burke, McKeen, & McKenna, 1993). If gender equity just centers on increasing access to technical skills, “concepts and tools will be misunderstood and ineffective” (Porter and Smyth, 1999, 332).

However, a teacher-student gender match alone cannot address systemic inequity, especially without taking intersectionality of identity into account (Rezai-Rashti & Goli, 2010). For example, this paper explores the social construct of gender in its binary form, and we understand that different contexts may change these definitions geographically and temporally.

### **Limitations**

We understand that in order to draw causal inferences, we must sacrifice some amount of external validity during the sample restriction process. We only include 63 schools in the analytic samples to ensure that sample students were randomly assigned teachers upon their entry to middle school in 2013-14 and did not sort in or out of their teachers’ class by 2014-15. This leads to systematic differences between schools included (N=63) and excluded (N=49) in these analyses. Schools in this

analysis appear to be more likely located in economically developed and urban areas, serve a better educated population, and have significantly smaller class sizes compared to schools excluded. As a result, these findings are not generalizable to schools in disadvantaged areas; these schools are more likely to fail to restrictively implement random assignments. Excluding these schools helps preserve the internal validity of this study. Future studies should look more closely at the assignment of teachers and students in schools from remote areas and if possible, conduct researcher-designed experiments to test the robustness of these findings across different school types.

Additionally, due to the self-report nature of CEPS data, there are potentially large measurement errors embedded in the key variables. For instance, the subject confidence variable only captures students' response to a single survey item therefore is potentially not accurately capturing the latent construct of students' confidence in learning. Great caution is suggested in interpreting these results and generalizing these findings to common practice. Future research should be focused on developing measures that have better validity and reliability and using them to estimate the advisor effects.

### References

- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics* 33, no. 1 (January 2015): 63–89. <https://doi.org/10.1086/677391>.
- Aucejo, Esteban M., Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter. "Teachers and the Gender Gap in Reading Achievement." *Journal of Human Capital* 16, no. 3 (September 1, 2022): 372–403. <https://doi.org/10.1086/719731>.
- Bettinger, Eric P, and Bridget Terry Long. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review* 95, no. 2 (April 1, 2005): 152–57. <https://doi.org/10.1257/000282805774670149>.
- Carrell, Scott E., Marianne E. Page, and James E. West. "Sex and Science: How Professor Gender Perpetuates the Gender Gap \*." *Quarterly Journal of Economics* 125, no. 3 (August 2010): 1101–44. <https://doi.org/10.1162/qjec.2010.125.3.1101>.
- Cho, Insook. "The Effect of Teacher–Student Gender Matching: Evidence from OECD Countries." *Economics of Education Review* 31, no. 3 (June 2012): 54–67. <https://doi.org/10.1016/j.econedurev.2012.02.002>.
- Cochran-Smith, Marilyn, Ana Maria Villegas, Linda Abrams, Laura Chavez-Moreno, Tammy Mills, and Rebecca Stern. "Critiquing Teacher Preparation Research: An Overview of the Field, Part II." *Journal of Teacher Education* 66, no. 2 (March 2015): 109–21. <https://doi.org/10.1177/0022487114558268>.
- Dee, Thomas S. "A Teacher like Me: Does Race, Ethnicity, or Gender Matter?," 2023.
- . "Teachers and the Gender Gaps in Student Achievement," n.d.

- Driessen, Geert, and Annemarie van Langen. "Gender Differences in Primary and Secondary Education: Are Girls Really Outperforming Boys?" *International Review of Education* 59, no. 1 (June 2013): 67–86. <https://doi.org/10.1007/s11159-013-9352-6>.
- Eble, Alex, and Feng Hu. "Child Beliefs, Societal Beliefs, and Teacher-Student Identity Match." *Economics of Education Review* 77 (August 2020): 101994. <https://doi.org/10.1016/j.econedurev.2020.101994>.
- . "Role Models, the Formation of Beliefs, and Girls' Math Ability: Evidence from Random Assignment of Students in Chinese Middle Schools," n.d.
- Fernandez Castela, Carolin, and Birgit Kröner-Herwig. "Developmental Trajectories and Predictors of Externalizing Behavior: A Comparison of Girls and Boys." *Journal of Youth and Adolescence* 43, no. 5 (May 2014): 775–89. <https://doi.org/10.1007/s10964-013-0011-9>.
- Flore, Paulette C., and Jelte M. Wicherts. "Does Stereotype Threat Influence Performance of Girls in Stereotyped Domains? A Meta-Analysis." *Journal of School Psychology* 53, no. 1 (February 2015): 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>.
- Gong, Jie, Yi Lu, and Hong Song. "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes," n.d.
- Guan, Alice, Marilyn Thomas, Eric Vittinghoff, Lisa Bowleg, Christina Mangurian, and Paul Wesson. "An Investigation of Quantitative Methods for Assessing Intersectionality in Health Research: A Systematic Review." *SSM - Population Health* 16 (December 2021): 100977. <https://doi.org/10.1016/j.ssmph.2021.100977>.
- Hill, Heather C., Charalambos Y. Charalambous, and Mark J. Chin. "Teacher Characteristics and Student Learning in Mathematics: A Comprehensive Assessment." *Educational Policy* 33, no. 7 (November 2019): 1103–34. <https://doi.org/10.1177/0895904818755468>.



- Kulkarni, Tara, Amanda L. Sullivan, and Jiwon Kim. "Externalizing Behavior Problems and Low Academic Achievement: Does a Causal Relation Exist?" *Educational Psychology Review* 33, no. 3 (September 2021): 915–36. <https://doi.org/10.1007/s10648-020-09582-6>.
- Lavy, Victor. "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment." Cambridge, MA: National Bureau of Economic Research, August 2004. <https://doi.org/10.3386/w10678>.
- Lei, Hao, Xijing Wang, Ming Ming Chiu, Mingfeng Du, and Tongwei Xie. "Teacher-Student Relationship and Academic Achievement in China: Evidence from a Three-Level Meta-Analysis." *School Psychology International* 44, no. 1 (February 2023): 68–101. <https://doi.org/10.1177/01430343221122453>.
- Lim, Jaegeum, and Jonathan Meer. "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea." Cambridge, MA: National Bureau of Economic Research, July 2015. <https://doi.org/10.3386/w21407>.
- Lockwood, Penelope, Tara C. Marshall, and Pamela Sadler. "Promoting Success or Preventing Failure: Cultural Differences in Motivation by Positive and Negative Role Models." *Personality and Social Psychology Bulletin* 31, no. 3 (March 2005): 379–92. <https://doi.org/10.1177/0146167204271598>.
- McIntyre, Rusty B, Charles G Lord, Dana M Gresky, Laura L Ten Eyck, G D Jay Frye, and Charles F Bond. "A SOCIAL IMPACT TREND IN THE EFFECTS OF ROLE MODELS ON ALLEVIATING WOMEN'S MATHEMATICS STEREOTYPE THREAT" 10, no. 9 (n.d.).
- Morgenroth, Thekla, Michelle K. Ryan, and Kim Peters. "The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals." *Review of General Psychology* 19, no. 4 (December 2015): 465–83. <https://doi.org/10.1037/gpr0000059>.

Paredes, Valentina. "A Teacher like Me or a Student like Me? Role Model versus Teacher Bias Effect."

*Economics of Education Review* 39 (April 2014): 38–49.

<https://doi.org/10.1016/j.econedurev.2013.12.001>.

———. "Mixed but Not Scrambled Gender Gaps in Single-Sex Classrooms," n.d.

Rezai-Rashti, Goli M., and Wayne J. Martino. "Black Male Teachers as Role Models: Resisting the

Homogenizing Impulse of Gender and Racial Affiliation." *American Educational Research Journal*

47, no. 1 (March 2010): 37–64. <https://doi.org/10.3102/0002831209351563>.

Rockoff, Jonah E. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel

Data." *American Economic Review* 94, no. 2 (April 1, 2004): 247–52.

<https://doi.org/10.1257/0002828041302244>.

Sansone, Dario. "Why Does Teacher Gender Matter?" *Economics of Education Review* 61 (December

2017): 9–18. <https://doi.org/10.1016/j.econedurev.2017.09.004>.

Spencer, S.J., Steele, C. M., and Quinn, D.M. Stereotype Threat and Women's Math Performance. *Journal*

of Experimental Social Psychology, 35 (1999), pp. 4-28, [10.1006/jesp.1998.1373](https://doi.org/10.1006/jesp.1998.1373)

Steinberg, Matthew P., and Rachel Garrett. "Classroom Composition and Measured Teacher

Performance: What Do Teacher Observation Scores Really Measure?" *Educational Evaluation*

and Policy Analysis 38, no. 2 (June 2016): 293–317. <https://doi.org/10.3102/0162373715616249>.

Su, Aoxue, Shuya Wan, Wei He, and Lianchun Dong. "Effect of Intelligence Mindsets on Math

Achievement for Chinese Primary School Students: Math Self-Efficacy and Failure Beliefs as

Mediators." *Frontiers in Psychology* 12 (March 26, 2021): 640349.

<https://doi.org/10.3389/fpsyg.2021.640349>.

Warren, Hannah. "Using Gender-Analysis Frameworks: Theoretical and Practical Reflections." *Gender &*

*Development* 15, no. 2 (July 2007): 187–98. <https://doi.org/10.1080/13552070701391847>.

Winters, Marcus A., Robert C. Haight, Thomas T. Swaim, and Katarzyna A. Pickering. "The Effect of Same-Gender Teacher Assignment on Student Achievement in the Elementary and Secondary Grades: Evidence from Panel Data." *Economics of Education Review* 34 (June 2013): 69–75.

<https://doi.org/10.1016/j.econedurev.2013.01.007>.

Xu, Di, and Qiujie Li. "Gender Achievement Gaps among Chinese Middle School Students and the Role of Teachers' Gender." *Economics of Education Review* 67 (December 2018): 82–93.

<https://doi.org/10.1016/j.econedurev.2018.10.002>.

Appendix

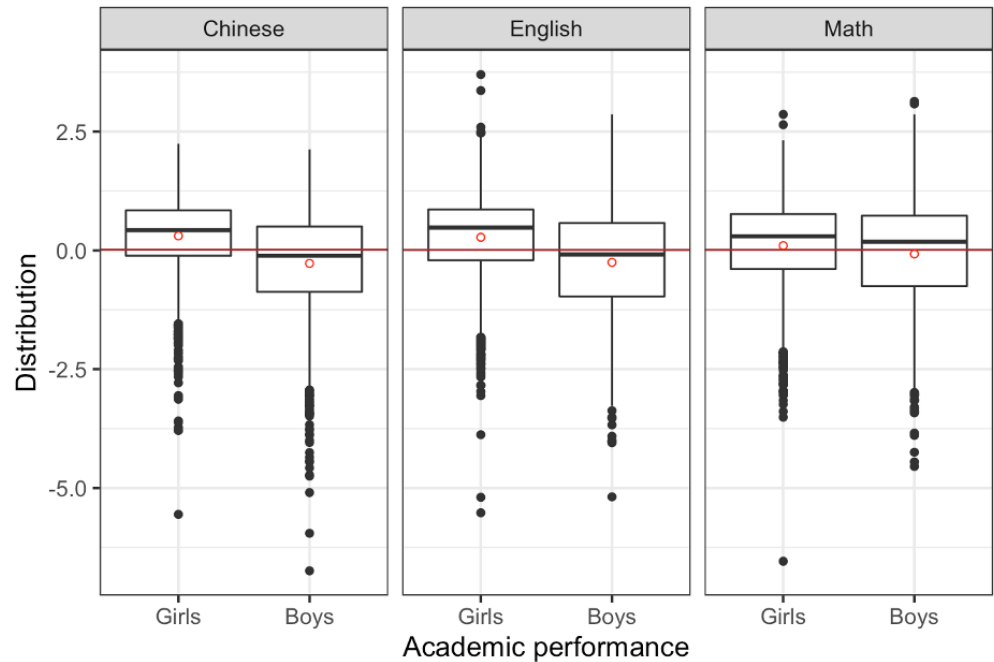


Figure 1. Distribution of student academic performance by student gender

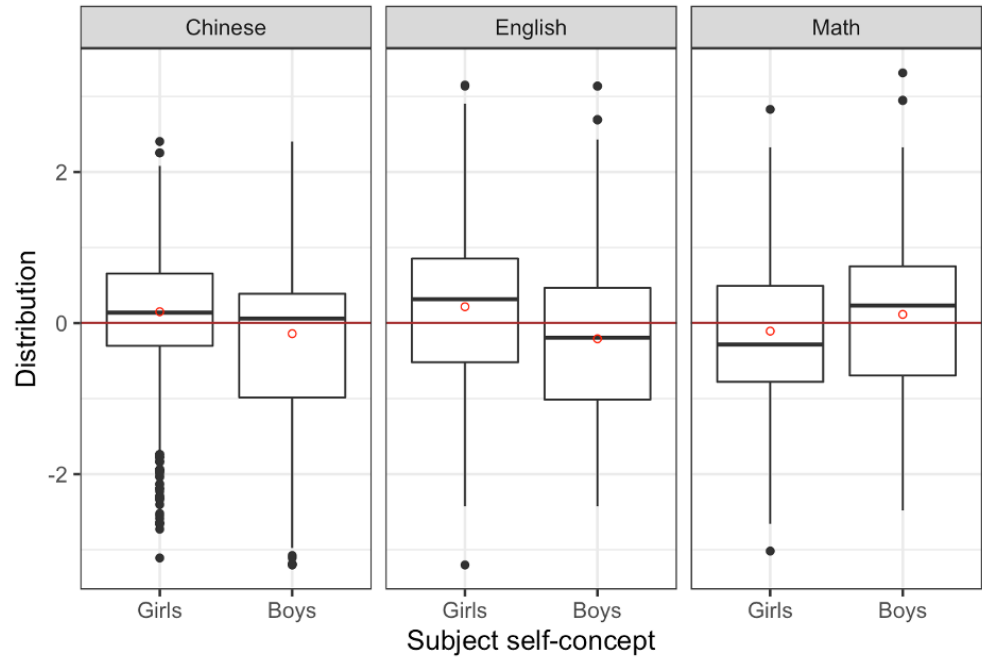


Figure 2. Distribution of subject self-concept by student gender

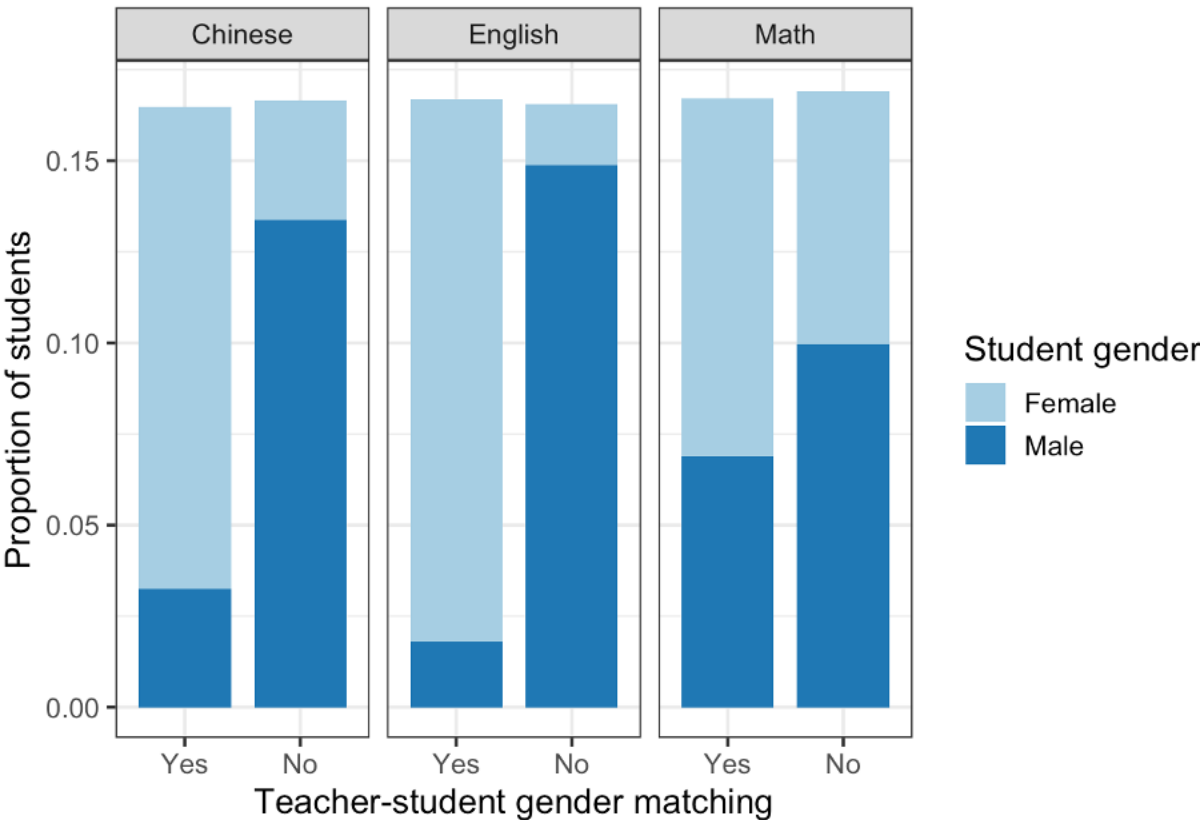


Figure 3. Teacher-student gender matching and student gender

Table 1. Analytic sample summary statistics

## Panel A. Chinese sample

	Female students, N = 2,464	Male students, N = 2,481
Match	80.03%	19.59%
Female teacher	80.03%	80.41%
Score	0.305 (0.814)	-0.273 (1.045)
Self-concept	0.150 (0.921)	-0.140 (1.036)

## Panel B. English sample

	Female students, N = 2,473	Male students, N = 2,489
Match	89.77%	10.85%
Female teacher	89.77%	89.15%
Score	0.276 (0.858)	-0.252 (1.044)
Self-concept	0.217 (0.935)	-0.208 (1.003)

## Panel C. Math sample

	Female students, N = 2,498	Male students, N = 2,518
Match	58.61%	40.87%
Female teacher	58.61%	59.13%
Score	0.100 (0.927)	-0.074 (1.042)
Self-concept	-0.106 (0.932)	0.115 (1.036)

*\*Cells display (%); mean (SD)*

Table 2. Assumption 1 check: observed balance between female and male teachers

## Panel A. Chinese sample

Teacher Characteristics	Female, N = 89	Male, N = 24	p-value
Age	37 (7)	41 (8)	0.018
Teaching experience (years)	15 (8)	19 (9)	0.015
Education attainment (years)	16 (1)	16 (1)	0.011
Teacher-advisor	28%	25%	0.8
Professional rank			0.2
Novice teacher	3.40%	4.20%	
Intermediate teacher	38%	17%	
Advanced teacher	37%	54%	
Senior teacher	21%	25%	

## Panel B. English sample

Teacher Characteristics	Female, N = 96	Male, N = 12	p-value
Age	38 (7)	40 (8)	0.5
Teaching experience (years)	16 (9)	18 (9)	0.4
Education attainment (years)	16 (1)	16 (1)	0.7
Teacher-advisor	29%	42%	0.5
Professional rank			0.8
Novice teacher	5.20%	0%	
Intermediate teacher	32%	25%	
Advanced teacher	43%	42%	
Senior teacher	20%	33%	

## Panel C. Math sample

Teacher Characteristics	Female, N = 65	Male, N = 51	p-value
Age	38 (8)	42 (7)	0.02
Teaching experience (years)	15 (8)	20 (9)	0.006
Education attainment (years)	16 (1)	16 (1)	0.002
Teacher-advisor	31%	31%	>0.9
Professional rank			0.6
Novice teacher	4.60%	2.00%	
Intermediate teacher	22%	14%	
Advanced teacher	48%	53%	
Senior teacher	26%	31%	

Table 3. Assumption 2 check: observed balance between students taught by female versus male teachers

	Chinese Sample		English Sample		Math Sample	
	Female teacher	Matched	Female teacher	Matched	Female teacher	Gender match
Baseline Chinese	0.016 (0.010)	0.018 (0.013)	-0.018 (0.009)	0.001 (0.008)	-0.026 (0.013)	0.025* (0.012)
Baseline English	-0.002 (0.015)	-0.023* (0.009)	0.004 (0.010)	-0.003 (0.009)	0.003 (0.017)	-0.003 (0.013)
Baseline math	-0.0003 (0.014)	0.003 (0.010)	0.013 (0.011)	0.003 (0.006)	0.033* (0.015)	-0.004 (0.012)
Baseline cognitive	-0.008 (0.012)	0.011 (0.011)	-0.011 (0.009)	0.0005 (0.005)	-0.047* (0.019)	-0.003 (0.012)
Female student	-0.016 (0.010)	0.606* (0.083)	0.005 (0.006)	0.791*** (0.064)	0.002 (0.010)	0.164 (0.085)
Age	0.011 (0.013)	0.003 (0.013)	0.001 (0.010)	-0.006 (0.009)	-0.001 (0.014)	-0.011 (0.014)
Only child	-0.013 (0.009)	0.003 (0.018)	0.001 (0.007)	0.008 (0.016)	-0.001 (0.012)	-0.007 (0.023)
Rural residency	-0.007 (0.008)	-0.020 (0.015)	-0.002 (0.007)	0.004 (0.014)	0.014 (0.016)	0.003 (0.020)
Migrant worker family	-0.017 (0.010)	-0.006 (0.016)	0.011 (0.008)	0.014 (0.012)	0.006 (0.018)	0.003 (0.025)
Mother education, years	-0.002 (0.001)	-0.002 (0.003)	-0.001 (0.001)	-0.002 (0.002)	0.0002 (0.002)	0.004 (0.003)
Father education, years	-0.001 (0.001)	-0.003 (0.002)	0.001 (0.001)	0.002 (0.002)	0.001 (0.002)	0.001 (0.003)
Family income	0.014 (0.011)	-0.004 (0.013)	0.019* (0.007)	-0.009 (0.010)	0.031* (0.011)	-0.017* (0.017)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
School-clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
F-Statistics	1.281 (df = 12; 62)	15.874*** (df = 12; 62)	0.806 (df = 12; 62)	35.06*** (df = 12; 62)	2.51** (df = 12; 62)	1.71 (df = 12; 62)
Observations	4,945	4,945	4,962	4,962	5,016	5,016
R <sup>2</sup>	0.615	0.547	0.647	0.772	0.622	0.539

Notes: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where the predictor variables (teacher education and experience) is regressed on baseline student prior score measures and characteristics. All models control for school fixed effects and cluster standard errors at school level.



Table 4. Gender gap in achievement

## Panel A. Baseline gender achievement gap

	Chinese	Chinese	Chinese	English	English	English	Math	Math	Math
	0.580**	0.590**	0.313**	0.513*	0.525*	0.241*	0.125*	0.139*	-0.246*
	*	*	*	**	**	**	**	**	**
Female student	-0.031	-0.03	-0.025	-0.031	-0.028	-0.024	-0.033	-0.029	-0.028
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cognitive score	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Other subject scores	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,945	4,945	4,945	4,962	4,962	4,962	5,016	5,016	5,016
R2	0.107	0.207	0.516	0.099	0.21	0.587	0.025	0.185	0.503

## Panel B. Wave 2 gender achievement gap

	Chinese	Chinese	Chinese	English	English	English	Math	Math	Math
	0.573**	0.172**	0.175**	0.522*	0.100*	0.108*	0.162*	0.069*	-0.078*
	*	*	*	**	**	**	**	*	**
Female student	-0.032	-0.022	-0.022	-0.032	-0.022	-0.024	-0.036	-0.023	-0.02
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline same subject score	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Baseline other subject scores	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,945	4,945	4,945	4,962	4,962	4,962	5,016	5,016	5,016
R2	0.106	0.527	0.605	0.096	0.681	0.705	0.027	0.557	0.612

## Panel C. Wave 2 gender gap in subject self-confidence

	Chinese	Chinese	Chinese	English	English	English	Math	Math	Math
	0.296**	0.183**	0.165**	0.422*	0.201*	0.173*	-0.229*	-0.286*	-0.287*
	*	*	*	**	**	**	**	**	**
Female student	-0.027	-0.027	-0.028	-0.037	-0.039	-0.04	-0.037	-0.032	-0.03
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline same subject score	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Baseline other subject scores	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,945	4,945	4,945	4,962	4,962	4,962	5,016	5,016	5,016
R2	0.037	0.07	0.076	0.078	0.239	0.247	0.031	0.228	0.232

Table 5. Teacher gender effects

Impact of Having a Female Teacher, by Subject

	Chinese Sample			English Sample			Math Sample		
Female	0.185** * (0.022)	0.174** * (0.021)	0.176** * (0.022)	0.098** * (0.023)	0.089** * (0.024)	0.091** * (0.023)	-0.075* **(0.019)	-0.083* **(0.025)	-0.081* **(0.025)
Student Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeroom Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,945	4,945	4,945	4,962	4,962	4,962	5,016	5,016	5,016
R <sup>2</sup>	0.608	0.611	0.611	0.710	0.711	0.712	0.615	0.617	0.620
Adjusted R <sup>2</sup>	0.601	0.603	0.603	0.706	0.706	0.707	0.609	0.610	0.613
Residual SE	0.619 (df = 4862)	0.618 (df = 4853)	0.618 (df = 4849)	0.538 (df = 4879)	0.538 (df = 4870)	0.537 (df = 4866)	0.619 (df = 4933)	0.618 (df = 4924)	0.616 (df = 4920)

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses. For each sample, the outcome variable is estimated three times: the first model controls for student covariates that include cubic polynomial functions of baseline Chinese, English, math, and CEPS cognitive test scores, and student characteristics; the second model adds homeroom covariates that include homeroom size and averaged student characteristics at homeroom level; the third model further adds teacher gender and homeroom advisor status. All models also control for school fixed effects and cluster standard errors at school level.



Homeroom Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,473	2,473	2,473	2,489	2,489	2,489	4,962	4,962	4,962
R <sup>2</sup>	0.688	0.689	0.691	0.703	0.704	0.706	0.710	0.711	0.712
Adjusted R <sup>2</sup>	0.678	0.678	0.679	0.693	0.693	0.694	0.705	0.706	0.707
Residual SE	0.487 (df = 2390)	0.487 (df = 2381)	0.486 (df = 2378)	0.578 (df = 2406)	0.578 (df = 2397)	0.577 (df = 2394)	0.538 (df = 4879)	0.538 (df = 4870)	0.537 (df = 4867)

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses.

Panel C. Math sample

	Female students			Male students			All students		
Match	0.165** (0.060)	0.181** (0.067)	0.178** (0.065)	-0.011 (0.045)	0.003 (0.048)	-0.005 (0.051)	0.050* (0.024)	0.051* (0.025)	0.051* (0.025)
Student Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeroom Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,498	2,498	2,498	2,518	2,518	2,518	5,016	5,016	5,016
R <sup>2</sup>	0.606	0.611	0.613	0.638	0.640	0.642	0.615	0.616	0.619

Adjusted R <sup>2</sup>	0.593	0.596	0.598	0.616	0.626	0.628	0.608	0.609	0.611
Residual SE	0.591	0.589	0.588	0.637	0.637	0.635	0.619	0.619	0.617
	(df =	(df =	(df =	(df =	(df =	(df =	(df =	(df =	(df =
	2415)	2406)	2403)	2435)	2426)	2423)	4933)	4924)	4921)

---

Note: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Cells report coefficients and associated standard errors in parentheses.

Table 7

Impact of Teacher-Student Gender Match on Subject Confidence, by Student Gender

## Panel A. Chinese sample

	Female students			Male students			All students		
Match	0.138 (0.115)	0.099 (0.123)	0.028 (0.100)	-0.090 (0.115)	-0.163 (0.116)	-0.184 (0.117)	0.117** * (0.031)	0.110** * (0.031)	0.113** * (0.031)
Student Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeroom Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,464	2,464	2,464	2,481	2,481	2,481	4,945	4,945	4,945
R <sup>2</sup>	0.079	0.087	0.096	0.073	0.079	0.080	0.077	0.082	0.085
Adjusted R <sup>2</sup>	0.048	0.052	0.060	0.042	0.044	0.044	0.061	0.065	0.068
Residual SE	0.899 (df = 2381)	0.897 (df = 2372)	0.0893 (df = 2369)	1.014 (df = 2398)	1.013 (df = 2389)	1.013 (df = 2386)	0.960 (df = 4862)	0.958 (df = 4853)	0.957 (df = 4850)

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses.

## Panel B. English sample

	Female students			Male students			All students		
Match	0.085 (0.133)	-0.033 (0.111)	-0.012 (0.095)	0.016 (0.104)	0.062 (0.087)	-0.019 (0.093)	0.130** (0.039)	0.132** (0.040)	0.130** (0.039)
Student Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeroom	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Covariates									
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,473	2,473	2,473	2,489	2,489	2,489	4,962	4,962	4,962
R <sup>2</sup>	0.247	0.255	0.260	0.259	0.264	0.268	0.262	0.266	0.269
Adjusted R <sup>2</sup>	0.221	0.227	0.231	0.233	0.236	0.239	0.249	0.252	0.255
Residual SE	0.825 (df = 2390)	0.822 (df = 2381)	0.820 (df = 2378)	0.879 (df = 2406)	0.877 (df = 2397)	0.875 (df = 2394)	0.860 (df = 4879)	0.859 (df = 4870)	0.857 (df = 4867)

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses.

Panel C. Math sample

	Female students			Male students			All students		
Match	0.155* (0.072)	0.123 (0.085)	0.105 (0.074)	-0.037 (0.065)	0.010 (0.068)	-0.041 (0.070)	0.024 (0.037)	0.026 (0.037)	0.025 (0.037)
Student Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeroom Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Teacher Covariates	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,498	2,498	2,498	2,518	2,518	2,518	5,016	5,016	5,016
R <sup>2</sup>	0.267	0.272	0.277	0.233	0.237	0.242	0.222	0.225	0.228
Adjusted R <sup>2</sup>	0.242	0.245	0.249	0.207	0.208	0.213	0.209	0.211	0.213

Residual SE	0.812	0.810	0.808	0.922	0.921	0.919	0.882	0.881	0.879
	(df =	(df =	(df =	(df =	(df =	(df =	(df =	(df =	(df =
	2415)	2406)	2403)	2435)	2426)	2423)	4933)	4924)	4921)

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Cells report coefficients and associated standard errors in parentheses.