The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes

Jie Gong, *National University of Singapore*

Yi Lu, *Tsinghua University and National University of Singapore*

Hong Song, *School of Economics, Fudan University*

This paper examines the role of teacher gender in education production. We extend student outcomes from traditionally focused academic achievement to noncognitive outcomes. Using a representative survey of middle school students in China, we focus on schools where student-teacher assignments are random. Our results show that having a female teacher raises girls' test scores and improves their mental status and social acclimation relative to those of boys. There is evidence that female teachers provide feedback differently to girls and boys and that having a female teacher alters girls' beliefs about commonly held gender stereotypes and increases their motivation to learn.

I. Introduction

The past two decades have witnessed the reversal of a remarkable gender gap in education. In the United States and many other developed countries, girls outperform boys in reading scores, top grade point average distribution in high school, and college attendance (Campbell 2000; Vincent-Lancrin

Contact the corresponding author, Yi Lu, at justinly6@gmail.com. Information concerning access to the data used in this paper is available as supplementary material online.

[*Journal of Labor Economics*, 2018, vol. 36, no. 3]

0734-306X/2018/3603-0005$10.00
Submitted September 16, 2016; Accepted June 5, 2017; Electronically published April 16, 2018

2008; Fortin, Oreopoulos, and Phipps 2015). This international phenomenon of female dominance in academic achievements raises new questions about the causes of this disparity. Are the differences primarily biological or nurtured by social influences? Are certain educational inputs, from either parents or teachers, more suitable for girls than boys? For example, is it only by coincidence that females comprise the majority of teachers in elementary and secondary schools, the period when girls achieve more and behave better than boys?

To gain insight into this question, we examine whether teacher gender affects the learning of girls and boys at the middle school level. A growing number of studies have attempted to document the effect of teacher gender at the primary school level (e.g., Winters et al. 2013; Antecol, Eren, and Ozbeklik 2015), secondary school level (e.g., Ehrenberg, Goldhaber, and Brewer 1995; Nixon and Robinson 1999; Dee 2005, 2007; Winters et al. 2013), and college level (e.g., Canes and Rosen 1995; Rothstein 1995; Neumark and Gardecki 1998; Bettinger and Long 2005; Hoffmann and Oreopoulos 2009; Carrell, Page, and West 2010). Most of these studies tend to focus on the impact on student academic performance and find that having a female teacher improves female students' educational outcomes. One exception is Antecol, Eren, and Ozbeklik (2015), who find that, at the primary school level, having a female teacher lowers the math test scores of female students. In contrast to the rich literature on academic outcomes, relatively less research has examined the effects of teacher gender on students' noncognitive outcomes—for example, attitudes, preferences, and socioemotional factors. The mechanisms behind teacher-student gender interactions also have limited empirical support.

We attempt to fill these gaps in the literature. First, we extend student outcomes from traditionally focused academic achievements to noncognitive outcomes. Since Jencks (1979), a number of studies have documented the importance of noncognitive skills in explaining academic achievement, labor market success, and other significant life outcomes (Heckman and Rubinstein 2001; Flossmann, Piatek, and Wichert 2008; Bertrand and Pan 2013; Heckman, Pinto, and Savelyev 2013; Segal 2013). To understand the role played by school influence in noncognitive outcomes, we focus on the middle school period, as this is the age at which noncognitive skills are thought to develop and mature (Borghans et al. 2008; Heckman and Kautz 2013).

Our second contribution is to uncover the mechanisms by which gender interactions between teacher and student lead to differential outcomes for girls and boys. Do teachers have a preference for same-gender students so that they provide differential attention or responses to female and male students? Or, given the same teaching behaviors, do girls and boys react and respond to instructions in different ways? Understanding these mechanisms is important when designing policies to mitigate gender differences at school (Dee 2004, 2007; Carrell, Page, and West 2010), especially when it is difficult to change the gender composition of a school's instructors. Our research provides insight into this question by using unique data on teacher behavior and student beliefs to directly test how a teacher's behavior differs by student gender as well as how students' beliefs and motivations related to learning are influenced by their teacher's gender.

We use a nationally representative survey of middle school teachers and students in China and focus on schools in which the assignment of students to classrooms is random. This allows us to mitigate potential selection problems—for example, high-achieving or better-motivated female students are more likely to be assigned to female teachers. By comparing outcomes for girls and boys taught by female versus male teachers, we are able to estimate whether and how teacher gender affects female and male students' outcomes differently.

Our results show that having a female teacher has positive and significant effects on girls' academic and noncognitive outcomes relative to those of boys. On academic performance, our estimates confirm a reverse gender gap: girls outperform boys on test scores. What is immediately remarkable is how female teachers enlarge girls' performance lead: they raise girls' test scores by about 19.8% of a standard deviation relative to those of boys. Moreover, having a female teacher also improves girls' self-assessment of their learning. When taught by a male teacher, although girls' absolute grades are higher than those of boys, their self-assessed outcomes are not. Having a female teacher significantly increases girls' self-assessed scores relative to those of boys, and the magnitude of the effect is even larger than the effect on actual test scores.

The pattern for students' noncognitive outcomes is subtler. If taught by a male, girls are more likely to feel depressed, blue, or unhappy at school than boys. Female teachers can overturn the gender gap: they do not seem to affect boys' outcomes but significantly improve girls' mental status. In addition, we find strong evidence that female teachers improve girls' overall satisfaction and social acclimation with classmates.

Examining the mechanisms that drive these results, we find evidence for both teacher- and student-based channels. Specifically, we find that female teachers tend to ask more questions of, give more praise to, and make fewer critical comments toward girls. In addition, we find that the presence of a female teacher counters the perception that girls are not as strong in math and better motivates girls to study the subject. Overall, our findings support both the stereotype threat hypothesis and role model theories.

In the last part of our empirical analysis, we examine whether students from lower socioeconomic backgrounds are more sensitive to the influence of teacher gender. Our findings suggest that when the mother is less educated, the student is an ethnic minority, or the parents migrated out and left the student behind, teacher gender has a much stronger impact on students' test scores. These patterns are consistent with the premise that disadvantaged students receive less parental investment, so school-based influences play a more important role in their skill development.

Our findings provide new evidence that teacher gender has a significant influence on student outcomes at the middle school level. This reinforces previous findings on the impact of teacher gender on student grades and also contributes to the literature on the effect on middle school students in particular. Some studies have found that having a female teacher positively affects female middle school students' achievement (Nixon and Robinson 1999; Dee 2007; Winters et al. 2013), while others have found no effect (Ehrenberg, Goldhaber, and Brewer 1995). Our study also extends previous research by using a setting with randomized student assignment. To our knowledge, only Dee (2004), Carrell, Page, and West (2010), and Antecol, Eren, and Ozbeklik (2015) use randomized or experimental data to account for endogenous assignment; their data sources, respectively, are Project STAR, the US Air Force Academy, and the National Evaluation of Teach for America.[1] Our estimates are based on randomized student assignments and a more representative sample of middle school students and teachers and therefore are arguably more general from a policy perspective.

More broadly, we also contribute to the literature on how school environment influences noncognitive skills. Previous studies highlight the central role played by school choice and school social networks. For example, Cullen, Jacob, and Levitt (2006) find that gaining access to sought-after public schools improves students' behavioral outcomes, as evidenced by a lower level of disciplinary incidents and arrest rates. Angrist, Bettinger, and Kremer (2006) find that winners of vouchers for private schools in Colombia were more likely to finish the eighth grade, worked less, and were less likely to marry or cohabit as teenagers than those who did not win vouchers. Lavy (2010) evaluates a program of free choice among public schools in Tel Aviv, Israel, and finds better student behavioral outcomes, such as more positive teacher-student relationships, better social acclimation and satisfaction at school, and less violence and disruption in the classroom. Lavy and Sand (2015) exploit conditional random assignment in middle schools in Tel Aviv and find that social networks have positive effects on students' noncognitive behavioral outcomes: greater cooperation, reduction in violent behavior, and improvements in social satisfaction in class. Our study complements the literature by providing evidence of how one component of the educational experience, teacher gender, influences student noncognitive outcomes. An important advantage of our study is that our findings are not limited to a particular type of school—that is, public or private—because the estimates stem from within-school comparisons.

[1] Several studies deal with the selection problem by making within-student and within-teacher comparisons. See, e.g., Dee (2007), Hoffmann and Oreopoulos (2009), and Fairlie, Hoffmann, and Oreopoulos (2014).

II. Data and Variables

In the Chinese educational system, middle school students are assigned to classrooms at the beginning of the seventh grade and take the same courses throughout their 3 years of middle school. Students are required to take three core subjects—Chinese, mathematics, and English—and a set of subsidiary subjects. During a regular school day, students remain in the same classroom all day, and different teachers come to the classroom to deliver subject-specific lectures. A head teacher, who is usually one of the core-subject teachers, oversees class activities and individual student progress for a given classroom. While subject teachers ensure that students achieve their learning goals, the head teacher is also responsible for all class matters and students' social lives—for example, setting seating plans, organizing extracurricular events, and overseeing student discipline. In addition, the head teacher regularly gives feedback to students and their parents regarding academic performance and behavior.

We obtained data from the 2014 China Education Panel Survey (CEPS), a nationally representative survey that covers middle schools from 28 counties and city districts.[2] Our sample includes 8,988 students in the seventh and ninth grades across 208 distinct classrooms and 67 schools. Data for each student's demographic characteristics are collected from student questionnaires. Table 1 reports the summary statistics for our main outcome and control variables.

Our academic performance measures are student exam scores in the core subjects provided by their respective school administration offices. These subjects—Chinese, mathematics, and English—are compulsory and are the main components in the standard tests for admission to senior high school. Within a school, teachers who teach the same subject use the same syllabus and give the same exams during a common testing period. Test scores in the core subjects are therefore a consistent measure of academic achievement across students from the same grade in the same school.[3] We supplement our test score measures with self-assessed performance scores collected from the student questionnaire, on which students were asked to report whether they have difficulty in learning each subject on a scale from 1 (a lot) to 4 (not at all).

[2] CEPS is the first and largest nationally representative education survey in China. Officially started in 2013, the survey applies a stratified sampling design in which four middle schools and four classrooms within each school are chosen to represent a given county or city district.

[3] Exams are graded in a rigorous and consistent manner with each student's name, class, and ID information hidden from the grader. Within a grade at the same school, teachers divide the grading work so that the same question is typically graded by the same teacher using a consistent standard.

Table 1
Summary Statistics

	Mean (1)	SD (2)	Observations (3)
A. Outcome variable:			
Academic outcomes:			
Test score	81.26	28.34	25,783
Self-assessment score	2.48	.91	26,301
Noncognitive outcomes:			
Depressed	2.24	1.00	8,772
Blue	1.98	1.06	8,743
Unhappy	2.28	1.05	8,762
Pessimistic	1.75	1.07	8,734
School life is fulfilling	3.38	.86	8,852
Confident about future	3.26	.72	8,924
Social activities: public enrichment	2.02	1.04	8,686
Social activities: private recreation	2.44	1.28	8,653
B. Regressor of interests:			
Female head teacher	.65	.48	8,988
Female subject teacher	.75	.43	26,211
Female student	.49	.50	8,910
C. Predetermined variables:			
Student age	13.94	1.35	8,815
Minority	.11	.31	8,968
Local residence	.80	.40	8,811
Rural residence	.45	.50	8,510
Only child in family	.51	.50	8,986
Attend kindergarten	.82	.39	8,912
Repeat grade in primary school	.11	.32	8,988
Skip grade in primary school	.02	.12	8,966
Academic ranking in primary school	15.65	11.82	8,121
Express opinions clearly in primary school	3.14	.82	8,619
Respond quickly in primary school	3.03	.79	8,625
Learn new stuff quickly in primary school	3.02	.82	8,559
Mother's education	9.98	3.60	8,966
Father's education	10.45	3.29	8,966

To measure students' noncognitive outcomes, we use student responses to eight questionnaire items. Specifically, four questions asked students about the frequency of the following feelings during the previous 7 days on a scale from 1 (never) to 5 (always): (1) depressed, (2) blue, (3) unhappy, or (4) that life is meaningless. Two questions asked students to rate how much they agree with the following statements on a scale from 1 (strongly disagree) to 4 (strongly agree): (5) the student feels that school life is rich and fulfilling and (6) the student feels confident about his or her future. Last, two questions asked students to rate how frequently they take part in the following activities on a scale from 1 (never) to 6 (always): (7) going to museums, zoos, or science parks with classmates from school and (8) going to movies, plays, or sports events with classmates from school. For regression analysis, the variables are normalized to have means of 0 and standard deviations of 1.

In addition to estimating the effects of teacher gender on each of these eight noncognitive variables, we follow an aggregation method proposed by Kling, Liebman, and Katz (2007) to understand the overall effects of teacher gender. Specifically, we first use the principal component analysis method to group the eight noncognitive variables into two categories, that is, the level of mental stress and the level of social acclimation and satisfaction (details are presented in app. A; apps. A, B are available online). We then calculate the average effect size (AES) for each category, which is a weighted average of the z-scores of its component.[4] This aggregation improves statistical power to detect effects that are consistent across specific outcomes when the outcomes also have idiosyncratic variation.

Our main teacher-level variable of interest is gender. In addition, we obtain information from teacher questionnaires about which subject each respondent taught and his or her age, marital status, years of schooling, credential status, experience, tenure status, and professional job title at the current school. Our summary statistics show that 64.9% of head teachers and 75.4% of subject teachers in our sample are female. Within each core subject, female teachers are also more common than male teachers: 75% in Chinese, 62.5% in math, and 89.5% in English.[5]

III. Estimation Strategy

A. Class Assignment and Regression Sample

Our research question concerns the effect of teacher gender on student outcomes. Understanding how students are matched to teachers and classrooms is therefore critical to our estimation and analysis. Middle schools in China use various methods to assign students. In some schools, students take placement exams prior to beginning their first academic year, and their scores are used to assign them to classrooms. In other schools, students are assigned on the basis of local residency.

More recently, an increasingly large number of schools have begun to employ random assignment to place students into classrooms. This approach is heavily promoted by the Ministry of Education as ensuring equal and fair opportunities for all students during their compulsory education years (through the ninth grade). Schools that use randomized assignments typically rely on a computer program that can incorporate information on class size, gender,

[4] Specifically, we define the AES of teacher gender on noncognitive category c as $\mathrm{AES}_c = (1/n_c)\Sigma_{n=1}^{n_c} e_{kc}/\sigma_{kc}$, where n_c is the number of outcomes in category c (e.g., four for the category of mental stress), e_{kc} is the estimated effect for outcome k of category c, and σ_{kc} is the standard deviation of the outcome variables.

[5] Our observed female dominance in secondary school teachers is a global phenomenon. See, e.g., Holmlund and Sund (2008) and Winters et al. (2013).

migrant status, and other dimensions to ensure proper balance in the randomization process. Alternatively, if enrollment is small and manageable, parents of incoming students are invited to draw lots to determine their child's class placement. In these cases, once student assignments have been determined the head and subject teachers also draw lots to determine which classrooms they will teach and manage.

In this study, we focus on schools in which students are randomly assigned to classrooms. We require that a school's classroom-assignment procedure meet three conditions: (i) the school principal reports that students are randomly assigned to classrooms; (ii) after students have been assigned to classrooms at the beginning of the seventh grade, the school does not rearrange their classes for grades 8 and 9; and (iii) all head teachers in the same grade report that students in the respective grade are not assigned by test scores.[6] Using our criteria, we find that 59.8% of the schools in the 2014 CEPS database assign classrooms randomly, translating into a sample of 8,988 students across 208 classrooms and 67 schools. Since each student in our sample is randomly matched to both a head teacher and subject teachers and stays in the same classroom for the next 3 years, our sample mitigates any potential concerns regarding self-selection of students to classrooms or teachers.[7]

To verify the randomness of classroom assignment for our sample, we conduct a balance test between classrooms with a female head teacher and those with a male head teacher using several baseline characteristics (i.e., values that are determined before the classroom assignment). If the assignment process is truly random, these two groups of students should be similar across these predetermined characteristics. To this end, we solicit a rich set of baseline variables in the survey. They include (1) students' demographic information: female (1 if a student is a female and 0 otherwise), student age, minority (1 if a student belongs to a minority ethnic group and 0 otherwise), local residence

[6] Criteria are based on reports in the teacher questionnaire. For the first condition, all school principals were asked to report which of the following assignment rules they use to place students: (*a*) preenrollment test scores, (*b*) students' residential status, (*c*) random assignment, or (*d*) other factors. We restrict our sample to schools that use rule *c*. Second, the same principals were asked whether their schools rearrange classrooms for grades 8 and 9; we exclude those that do so. Finally, each head teacher was asked whether students in the grade level taught are assigned by test scores; again, we drop the entire grade if any head teacher answered yes.

[7] One might worry that parents complain about classroom randomization outcomes, as in the STAR experiment, and demand that their children move to other classrooms. We do not have direct information in the survey data. However, school principals are asked whether parents requested assignment of their children to specific teachers' classes, on a scale of 1 (completely irrelevant), 2 (irrelevant), 3 (slightly relevant), or 4 (absolutely relevant). Among the 67 schools that pass the randomization restriction, only one school's principal reported a 4 (absolutely relevant), and another nine reported a 3 (slightly relevant). Excluding these 10 schools that may have reassigned students does not change our estimation results.

(1 if a student is local and 0 otherwise), and only child in family (1 if a student is the only child in the family and 0 otherwise); (2) students' baseline academic outcomes, proxied by their academic performance in kindergarten and primary schools: attend kindergarten (1 if a student attended kindergarten and 0 otherwise), repeat grade in primary school (1 if a student repeated a grade in primary school and 0 otherwise), skip grade in primary school (1 if a student skipped a grade in primary school and 0 otherwise), and academic ranking in primary school (ranking of academic performance [in class] in primary school); (3) students' baseline noncognitive outcomes: express opinions clearly in primary school (a self-reported score from 1 [disagree] to 5 [agree] on whether they expressed opinions clearly in primary school), respond quickly in primary school (a self-reported score from 1 [disagree] to 5 [agree] on whether they responded quickly in primary school), and learn new stuff quickly in primary school (a self-reported score from 1 [disagree] to 5 [agree] on whether they learned new material quickly in primary school); and (4) parents' information: mother's education (mother's years of schooling) and father's education (father's years of schooling).

For the balance test, we follow Antecol, Eren, and Ozbeklik (2015) and regress the female head teacher dummy on students' baseline characteristics, controlling for block fixed effects (i.e., same grade in the same school). Table 2, column 1, reports regression results. We observe that none of the 15 baseline characteristics is statistically and economically significant. An *F*-statistic shows that they are also jointly insignificant. These results suggest that the student characteristics in our sample are well balanced across classes managed by female versus male head teachers. In columns 2–4, we further investigate whether student characteristics are correlated with subject teachers' gender. We find that most of the coefficients are statistically insignificant, and all are small in magnitude. All of the *F*-statistics are also small and statistically insignificant. Taken together, these results indicate that randomization ensures that student characteristics and teacher gender are uncorrelated.

A limitation of our data set, compared with those of previous studies using randomized assignment (e.g., Carrell, Page, and West 2010; Antecol, Eren, and Ozbeklik 2015), is the lack of direct measures of baseline academic and noncognitive outcomes. Our proxies might be insufficient to capture students' baseline situations, and therefore the balance test may not lend full support to the randomization setting. To address this concern, we include all of the student-level predetermined characteristics in the main analyses to further improve the balance between the two types of classrooms as well as our estimation efficiency.

We also conduct two robustness checks on the validity of our research setting. If our sample includes nonrandomized classrooms, our estimated teacher gender effects would be biased. This is because some schools may declare that they randomized but in fact did not; however, we cannot identify such schools and eliminate them from the sample. In one exercise, we

Table 2
Balancing Test

	Female Head Teacher (1)	Female Math Teacher (2)	Female Chinese Teacher (3)	Female English Teacher (4)
Female student	.005	.002	−.003	.001
	(.006)	(.005)	(.003)	(.004)
Student age	−.007	−.005	−.004	.007*
	(.006)	(.006)	(.005)	(.004)
Minority	.003	.029*	−.010	.000
	(.019)	(.017)	(.011)	(.008)
Local residence	−.001	−.025	.025*	−.002
	(.010)	(.019)	(.014)	(.006)
Rural residence	−.015	−.014	.021*	−.006
	(.014)	(.015)	(.012)	(.007)
Only child in family	.004	.008	−.005	.001
	(.008)	(.009)	(.006)	(.004)
Attend kindergarten	.012	.015	−.022	−.005
	(.014)	(.016)	(.014)	(.007)
Repeat grade in primary school	−.023	−.011	.008	.004
	(.016)	(.019)	(.014)	(.006)
Skip grade in primary school	−.028	−.028	.015	.030
	(.023)	(.032)	(.020)	(.020)
Academic ranking in primary school	−.000	−.000	−.000	−.000
	(.001)	(.001)	(.000)	(.000)
Express opinions clearly in primary school	−.004	−.008	.004	−.001
	(.005)	(.006)	(.004)	(.002)
Respond quickly in primary school	.008	.006	.006	−.004
	(.005)	(.005)	(.004)	(.003)
Learn new stuff quickly in primary school	.004	.010**	.001	−.001
	(.004)	(.005)	(.002)	(.002)
Mother's education	.001	.002	.000	.001
	(.001)	(.002)	(.001)	(.001)
Father's education	.002	−.001	.000	−.000
	(.002)	(.002)	(.001)	(.001)
Test for joint significance:				
F-statistics	1.410	.990	1.180	.900
p-value	.154	.474	.299	.568
Block fixed effects	Yes	Yes	Yes	Yes
Observations	7,025	7,025	6,959	6,849
R^2	.629	.582	.791	.622

NOTE.—Each cell presents the coefficient and standard error (in parentheses) for the listed student predetermined variables from the regression in which the dependent variable is the teacher gender and the independent variables are all of the student predetermined characteristics. Standard errors are clustered at the block level.
* Significant at the 10% level.
** Significant at the 5% level.

estimate a sample of schools that fail to pass our randomization criteria—that is, schools that nonrandomly assigned students. By comparing those results to our baseline estimates (i.e., those using the randomized-classrooms sample), one can infer the direction of bias, if any. In the other test, we randomly drop schools from the regression sample. If the baseline sample mostly consists of randomized classrooms, results from the reduced sample should not change dramatically from the baseline estimates. The results support our random assignment setting. Details are presented in Section IV.D.

B. Estimation Framework

To estimate the effect of teacher gender on student academic and noncognitive outcomes, we use the following regression model:

$$\begin{aligned} Y_{icb} = {} & \beta_1 \text{Femstud}_{icb} + \beta_2 \text{Femteach}_{cb} + \beta_3 \text{Femteach}_{cb} \\ & \times \text{Femstud}_{icb} + W'_{icb}\phi + D_b + \varepsilon_{icb}, \end{aligned} \tag{1}$$

where Y_{icb} is the outcome measure for student i in classroom c of block b (i.e., classrooms in the same school and the same grade), Femstud_{icb} is an indicator for whether the student is female, Femteach_{cb} is an indicator for whether the teacher for classroom c of block b is female, W_{icb} is a set of student controls (i.e., the student characteristics used in the balance test in table 2) and teacher controls (i.e., age, marital status, years of schooling, whether they graduated from a normal college, credential status, experience in the profession, whether they had prior teaching experience in other schools, tenure status, professional job title [for subject teachers only], and whether they teach core subjects [for head teachers only]), D_b is the block fixed effects, and ε_{icb} is the error term.

The β coefficients are the main coefficients of interest: β_1 captures the difference in mean academic performance and noncognitive outcomes between female and male students when they have a male teacher, β_2 captures the value added by having a female teacher, and β_3 captures the effect of teacher gender on student gender differences. An unbiased estimation of the β coefficients requires that conditional on the controls, Femteach_{cb} and Femstud_{icb} are uncorrelated with the error term ε_{icb}. These identifying assumptions hold in a setting of random classroom assignment, which is verified in the previous section.

To estimate equation (1), we pool seventh and ninth grade together for a larger sample and better estimation precision. As randomization is conducted within a block (i.e., same grade and same school) and students may nonrandomly select their middle school,[8] we include block fixed effects D_b in the re-

[8] In most regions of China, middle school quality varies. Better schools generally provide better facilities and hire more competent teachers, and therefore they attract different types of students than less prestigious schools. As a result, comparison of

gressions to control for all block-level factors in our cross-sectional data that may influence students' school choice. Our identification, therefore, comes from the comparison between teachers and their students in the same block. Standard errors are clustered at the block level to account for any correlation in outcomes for students in the same block. In a robustness check, we estimate a specification with classroom fixed effects and standard errors clustered at the classroom level to explore differences in student and teacher gender within each classroom for identification. In a further robustness check, we estimate equation (1) separately for the seventh and ninth grades, in which we include school fixed effect and cluster standard errors at the school level.

IV. Main Results

A. Teacher Gender and Student Academic Performance

Table 3 presents our estimates of the teacher gender effect on student academic performance. For each specification, we report both the effects on test scores (cols. 1, 3, 5) and the effects on student self-assessed scores (cols. 2, 4, 6). Both performance measures are normalized by subject, grade, and school and therefore have means of 0 and standard deviations of 1. All regressions include subject fixed effects and block fixed effects. Similar to Dee (2007), we first show results that include only student controls in panel A and then results that also include teacher controls in panel B. Results are similar with or without the inclusion of teacher controls, and therefore we focus on the results using full controls (panel B) when discussing magnitude.

As both the head teacher and the subject teacher might be relevant to students' academic achievement, we test the relevance of the head teacher's gender and the subject teacher's gender separately. We observe that while the coefficients of female head teacher × female student are statistically and economically insignificant (cols. 1, 2), the gender interaction between students and subject teacher—female subject teacher × female student—are positive and statistically significant. The patterns remain robust when both head and subject teachers' gender are included simultaneously (cols. 5, 6). These results suggest that having a female subject teacher improves girls' academic performance relative to that of boys, but no such effects are found for having a female head teacher.

The differential gender effects of head and subject teachers are not surprising and reflect their distinct roles in the classroom. Subject teachers teach a specific course and ensure that students achieve the associated learning goals. They are responsible for—and are often evaluated on the basis of— students' performance in the specific subject. Head teachers perform a more manage-

students from different schools may contaminate our estimates with differences in student characteristics and learning environments among schools. In contrast, within the same school and same grade, the random assignment of students to teachers ensures that student characteristics are uncorrelated with other confounders.

rial role; they oversee class activities and student progress for the whole classroom. Rather than focusing on performance in a particular course, head teachers likely expend effort cultivating students' general skills, behaviors, and disciplines. We posit that head teachers may more likely influence students' noncognitive skills.

In terms of economic magnitude, our estimates from the full specifications (cols. 5, 6) imply that having a female subject teacher raises girls' test scores by 19.8% of a standard deviation and their self-assessment by 31.6% of a standard deviation relative to those of boys. Our findings are consistent with existing evidence on teacher-student gender dynamics in middle school and above. More specifically, at the middle school level our findings confirm the results of Dee (2007) and Winters et al. (2013), who show that assignment to a female teacher significantly improves girls' test scores relative to those of boys. Our estimates may seem high relative to the effects they document;[9] this could partly be due to the differences in the sample we consider and to the randomized setting of our data. Specifically, we examine middle schools in China, the economy and education system of which are different from—and plausibly less developed than—the US schools Dee (2007) and Winters et al. (2013) study. In addition, we use random assignment and control for block or classroom fixed effects. The two studies cited above, on the other hand, used nonrandom data and acknowledge that the magnitudes of the effects in their analysis may potentially be biased. Further evidence to suggest this is provided by our robustness check in Section IV.D: when we repeat the analysis using nonrandomized classrooms, estimated effects are indeed smaller in magnitude.

The regression results also confirm girls' performance lead on test scores. The coefficients on the female student dummy are positive and significant, suggesting that girls achieve higher scores than boys (16.2% of a standard deviation) when taught by a male subject teacher. Female teachers further enlarge the gender difference (to 36% of a standard deviation). This is consistent with previous evidence of female dominance in academic performance.[10]

Another interesting finding concerns students' self-assessment scores. Although girls' absolute grades are higher than boys, their self-assessed outcomes

[9] Using conservative estimates, teacher gender effects—i.e., the increase in girls' test scores relative to that for boys when taught by a female teacher—are about 5.4% in Dee (2007) and 0.8% in Winters et al. (2013).

[10] A similar reverse gender gap has been found among students in the United States and Sweden. US educational statistics report that between 1973 and 1999, female students consistently dominated male students in reading scores, while the gender gap in mathematics decreased and became insignificant (Campbell 2000). The Swedish National Agency for Education reports that girls score significantly higher than boys on Swedish and English tests, and there is no obvious gender difference in mathematics scores (Holmlund and Sund 2008).

Table 3
Effect of Teacher Gender on Academic Outcomes

	Test Score (1)	Self-Assessment (2)	Test Score (3)	Self-Assessment (4)	Test Score (5)	Self-Assessment (6)
A. Without teacher controls:						
Female subject teacher ×						
female student			.183***	.289***	.200***	.316***
			(.052)	(.044)	(.051)	(.051)
Female subject teacher			−.036	−.125***	−.063**	−.146***
			(.032)	(.029)	(.031)	(.033)
Female head teacher ×						
female student	.024	.040			−.042	−.066*
	(.046)	(.034)			(.043)	(.037)
Female head teacher	.115*	.036			.141**	.087**
	(.060)	(.043)			(.059)	(.044)
Female student	.270***	.123***	.146***	−.071*	.160***	−.049
	(.037)	(.025)	(.047)	(.039)	(.050)	(.034)
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	No	No	No	No	No
Observations	18,202	18,601	18,202	18,601	18,202	18,601
R^2	.259	.080	.260	.083	.261	.084

B. With teacher controls:						
Female subject teacher ×						
female student			.182***	.289***	.198***	.316***
			(.051)	(.045)	(.050)	(.051)
Female subject teacher			−.035	−.126***	−.068**	−.151***
			(.030)	(.029)	(.028)	(.033)
Female head teacher ×						
female student	.024	.040			−.040	−.065*
	(.045)	(.034)			(.042)	(.037)
Female head teacher	.053	.059			.089	.115***
	(.063)	(.041)			(.063)	(.042)
Female student	.271***	.123***	.147***	−.071*	.162***	−.049
	(.037)	(.025)	(.046)	(.039)	(.049)	(.034)
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,202	18,601	18,202	18,601	18,202	18,601
R^2	.264	.081	.261	.084	.267	.086

NOTE.—Test score and self-assessment score are normalized by subject, grade, and school, to obtain a mean of 0 and a standard deviation of 1. Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

are not. This may reflect a gender gap in confidence about academic achievement. Female teachers help reverse the gap: they raise girls' self-assessment by 31.6% of a standard deviation relative to that of boys. Having a female teacher seems to have a greater effect on girls' self-assessment than it does on actual scores, suggesting that teacher gender plays an important role in students' confidence in academic achievements.

B. Teacher Gender and Students' Noncognitive Outcomes

To estimate the effects of the head teacher's gender on students' noncognitive outcomes, we normalize each of the eight noncognitive outcome variables to obtain a mean of 0 and a standard deviation of 1.[11] We also report the AES indices following Kling, Liebman, and Katz (2007) for the two categories, that is, mental stress as well as social acclimation and satisfaction.

Panel A of table 4 presents the estimated effects on students' mental stress. We present the results with block fixed effects and student controls (cols. 1–5) and the results that further include teacher controls (cols. 6–10). We consistently find that having a female teacher reduces girls' mental stress relative to that of boys. First, estimated coefficients on the female student dummy variables are mostly positive except for a negative but statistically insignificant coefficient for the likelihood of feeling pessimistic, indicating that girls are more likely to feel depressed, blue, or unhappy than boys when their head teacher is male. Second, estimates of the female teacher dummy variable are statistically and economically insignificant, suggesting that the head teacher's gender does not influence boys' mental stress levels. Finally and most importantly, estimated coefficients on the interaction term female teacher $\times$ female student are negative, indicating that having a female head teacher significantly reduces girls' mental stress relative to boys. Estimates are statistically significant for students' likelihood of feeling depressed, blue, unhappy, and the AES.

Panel B of table 4 shows the estimated effects on students' social acclimation and satisfaction. We again find positive effects of being assigned to a female head teacher on girls' confidence about the future and private recreation with peers. Estimated coefficients on the female student dummy variables are negative for confidence about the future and private recreation with peers and close to zero for the other two measures; this suggests that girls are less confident about the future and less likely to hang out with classmates when their head teacher is male. Coefficients on the female teacher dummy

[11] We focus on the head teacher's gender, as head teachers are more involved in students' activities at school than the subject teachers. In addition, a student's noncognitive outcomes do not vary by subject. Still, as a robustness check we include the ratio of female teachers in the classroom (including both head and subject teachers) as an alternative measure of teacher gender and report the results in Sec. IV.C and app. B.

variable are statistically and economically insignificant, suggesting that these outcomes for boys are not affected by the head teacher's gender. The key coefficients, female teacher × female student interaction, are positive and significant for students' confidence about the future and private recreation with peers, as is the AES. We interpret this as evidence that having a female head teacher improves girls' confidence and acclimation at school relative to those of boys.

Our findings across various specifications suggest that girls have lower noncognitive outcomes than boys—worse mental status, less confidence, and lower satisfaction at school—when they are assigned to male teachers. A female teacher can reduce, or even overturn, the gender gap by improving girls' noncognitive measures relative to those of boys.

C. Robustness Checks

We conduct multiple robustness checks: a specification with classroom fixed effects, a specification with student fixed effects, separate analyses by grade and by subject, controlling for test scores in the noncognitive outcome analyses, using alternative measures of teacher gender, and testing for sample attrition. Our findings remain robust, and we report the results in appendix B.

First, instead of block fixed effects, we estimate a specification that controls for classroom fixed effects and cluster standard errors at the classroom level. Tables B1 and B2 (tables A1–A3, B1–B12 are available online) present regression results for academic outcomes and noncognitive skills, respectively. Reassuringly, we find that all of our previous results hold in this additional specification, and the magnitudes remain similar.

Second, we test whether the findings are robust to the inclusion of student fixed effects. Notice that we can include student fixed effects in the academic performance regressions but not in the noncognitive skill regressions. This is because a student has academic outcomes in multiple subjects in the data, but there is no within-student variation in noncognitive measures. As shown in table B3, the estimated effect of teacher gender on students' academic outcomes remains significant and similar to baseline estimates.

Third, we estimate teacher gender effects separately for seventh and ninth graders. In doing so, we include only one grade from each school, and therefore block fixed effects collapse to school fixed effects and we cluster standard errors at the school level accordingly. Tables B4 and B5 report regression results for academic outcomes and noncognitive skills, respectively. We find that the effects on academic outcomes hold for both seventh and ninth graders, with the latter showing larger effects. For noncognitive outcomes, we find that results are significant only for the ninth graders, while estimates for the seventh graders are small and statistically insignificant. These results may reflect the importance of teacher-student interaction in that ninth graders

Table 4
Teacher Gender Effect on Noncognitive Outcomes: Mental Stress and Social Acclimation and Satisfaction

	Depressed (1)	Blue (2)	Unhappy (3)	Pessimistic (4)	AES (5)	Depressed (6)	Blue (7)	Unhappy (8)	Pessimistic (9)	AES (10)
A. Mental stress:										
Female teacher × female student	−.198***	−.173***	−.140**	−.024	−.134***	−.196***	−.191***	−.136**	−.029	−.138***
	(.057)	(.060)	(.056)	(.045)	(.045)	(.057)	(.057)	(.056)	(.045)	(.045)
Female teacher	.090	.023	.053	.007	.043	.086	.010	.055	.019	.043
	(.062)	(.051)	(.056)	(.066)	(.052)	(.062)	(.055)	(.058)	(.068)	(.053)
Female student	.287***	.087*	.153***	−.028	.125***	.285***	.104**	.151***	−.028	.128***
	(.044)	(.046)	(.041)	(.035)	(.033)	(.045)	(.042)	(.041)	(.036)	(.033)
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	6,895	6,895	6,895	6,895	6,895	6,620	6,620	6,620	6,620	6,620
R^2	.087	.084	.089	.066		.085	.084	.091	.067	

			Social Activities with Peers					Social Activities with Peers		
	School Life Is Fulfilling (1)	Confident about Future (2)	Public Enrichment (3)	Private Recreation (4)	AES (5)	School Life Is Fulfilling (6)	Confident about Future (7)	Public Enrichment (8)	Private Recreation (9)	AES (10)
B. Social acclimation and satisfaction:										
Female teacher × female student	.066	.108**	.029	.139***	.086***	.073	.099*	.018	.147***	.085***
	(.050)	(.054)	(.042)	(.047)	(.031)	(.052)	(.055)	(.042)	(.049)	(.032)
Female teacher	−.020	−.043	.081	−.012	.002	−.029	−.055	.059	.013	−.003
	(.068)	(.052)	(.051)	(.043)	(.036)	(.072)	(.060)	(.063)	(.055)	(.040)
Female student	.012	−.141***	.001	−.103**	−.058**	.010	−.130***	.001	−.113***	−.058**
	(.042)	(.048)	(.037)	(.040)	(.028)	(.044)	(.049)	(.038)	(.041)	(.029)
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	6,739	6,739	6,739	6,739	6,739	6,466	6,466	6,466	6,466	6,466
R^2	.100	.155	.244	.245		.102	.156	.249	.248	

NOTE.—Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level. AES = average effect size.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

have had more years with their teachers than seventh graders. One may expect that the stronger the "treatment"—that is, longer and more teacher-student interactions—the larger the impact teachers have on students' development and achievements.

Fourth, our main analysis of academic performance pools all three subjects together and estimates the average effects. Here, we estimate the effect separately for each subject. Estimation results are reported in table B6. We find significant effects for math and English but not for Chinese. One possible explanation is that female students have a strong comparative advantage in Chinese relative to male students, and having a female Chinese teacher may not further enlarge girls' lead.

Fifth, one concern regarding the effects on noncognitive outcomes is that they might be correlated with students' academic performance. For example, high achievers are more likely to feel confident and satisfied in school, whereas students who perform worse on academic tests are more prone to mental stress. If such correlations are strong, then a teacher's gender influences students' noncognitive outcomes primarily through the effects on academic performance. To separate the direct effect from the indirect effect via academic performance, we add students' test scores and their interactions with student gender and head teacher gender as controls for noncognitive outcomes. As shown in table B7, estimates remain almost the same after controlling for test scores. In other words, holding test scores constant, having a female teacher still improves the level of mental health, general satisfaction, and social acclimation for girls relative to boys.

Sixth, in our analysis of noncognitive outcomes, we focus on the head teacher's gender. To capture the scenario in which subject teachers may also influence students' development of noncognitive skills, we conduct two alternative specifications. First, we use the ratio of female teachers in the classroom (including both head and subject teachers) as an alternative measurement.[12] Estimation results are reported in table B8. We find similar results as in our baseline specification. Second, we include a dummy variable for female head teacher and the ratio of females among other subject teachers to separately test their importance. Estimation results are reported in table B9. While we find that our results for female head teacher are robust, coefficients on the interaction between female student and the ratio of female subject teachers are also significant, suggesting that having more females among the subject teachers also benefits girls' noncognitive outcomes relative to those of boys.

[12] We are unable to do this exercise for academic outcomes. This is because our regressor of interest for noncognitive outcomes is at the classroom level, while that for academic outcomes is at the classroom-subject level. Instead, in investigating academic outcomes, we separately estimate the roles played by head and subject teachers.

Seventh, there are missing values for student and teacher variables, and our estimates would be biased if teacher gender affects the likelihood of missing values. We regress an attrition dummy on teacher gender, student gender, and block fixed effects. As shown in table B10, coefficients on teacher gender are close to zero and not statistically significant; this indicates that our results are unlikely to be driven by teacher gender.

D. Further Checks on the Identifying Assumption

Our identification relies on the random assignment of students to classrooms, which ensures that student characteristics are uncorrelated with teacher gender. We start with a strong criterion to select the sample of randomized classrooms: the school principal and all head teachers in the same grade declare that random assignment is used. A balance test of student baseline characteristics also supports the identification assumption. However, one may still be concerned that the regression sample is contaminated by schools that reported using randomization but did not, resulting in biased estimates.

To further verify our identifying assumption, we conduct two empirical exercises. First, we estimate the effects of teacher gender among schools that failed our randomization restriction—that is, the principal or head teacher report nonrandom assignment. By comparing the results against our baseline estimates, we can infer the direction of bias—that is, if we indeed included nonrandomized classrooms in the baseline sample, we can determine whether our estimates are biased upward or downward. Table B11 reports the results for academic outcomes. Compared with the baseline findings that used a randomized sample, while the gender interaction between student and subject teacher (female subject teacher × female student) remains similar—the magnitude declines slightly—the effect of the head teacher's gender (female head teacher × female student) becomes negative and statistically significant. Table B12 reports the results for noncognitive outcomes. We find that the effects of the head teacher's gender (female head teacher × female student) become statistically insignificant and small in magnitude—substantially different from our baseline estimates.

These results suggest that when classrooms are not randomized and student composition differs, female teachers (including head and subject teachers), relative to male teachers, tend to be matched to classrooms in which female students are disadvantaged in both academic and cognitive outcomes relative to male students. As a result, estimates of the teacher gender effect would be biased downward if the sample includes nonrandomized classrooms.[13]

[13] Specifically, the estimator from the nonrandomization sample is $\text{plim}\hat{\beta}_3^{\text{non}} = \beta + \text{bias}$, whereas the estimator from the randomization sample is $\text{plim}\hat{\beta}_3^{\text{ran}} = \beta$. Hence, $\text{bias} = \hat{\beta}_3^{\text{non}} - \hat{\beta}_3^{\text{ran}} < 0$. In other words, if our regression sample includes schools with nonrandom classroom assignment, our estimates suggest lower bounds of the true effects.

For our second check, we randomly drop schools from the sample and see whether the regression results change dramatically. If our baseline sample contains mostly schools with randomized classrooms, estimates using the reduced sample should not seriously deviate from our baseline findings. To maintain sufficient sample size, we drop two schools each time and conduct a total of 2,211 (C^2_{67}) regressions for each outcome variable. Tables B1 and B2 plot the distribution of these estimated coefficients for academic outcomes and noncognitive skills, respectively. Clearly, all distributions are closely centered around the respective baseline estimates. The upper and lower bounds of each distribution also lie in the same direction as the baseline estimates. Together, these results suggest that our findings are unlikely to be severely biased by the possible inclusion of schools that nonrandomly assigned students.

V. Mechanism

Our main results in tables 3 and 4 suggest that relative to those of male students, female students' test scores and noncognitive outcomes are substantially better when they are taught by female teachers. In this section, we investigate possible mechanisms through which teacher gender affects female and male student outcomes differently.

A. Teacher Gender or Other Characteristics?

We start by examining whether our documented gender differences in student performance are primarily driven by teacher gender per se or by other teacher characteristics correlated with teacher gender. For example, students may respond differently to younger versus older teachers. To address the relevance of teachers' professional characteristics, we follow the approach used by Carrell, Page, and West (2010).[14] Specifically, we add the interactions between student/teacher gender dummies with the teacher-level controls that we included in the main specification. Including these interactions can identify whether certain teacher characteristics affect female and male students differently. As shown in table 5, we see that the estimate for teacher gender effect is similar to our baseline estimates. These results suggest that observed effects are driven primarily by a teacher's gender rather than by other teacher characteristics.

B. The Role Played by Teacher Gender: Teacher Behavior and Student Response

Given that teacher gender drives our results, we examine whether this effect is due to teacher behavior or student responses. Regarding teacher

[14] Carrell, Page, and West (2010) also analyze a teacher's value added to distinguish teacher gender effects from other teacher characteristic effects. However, this approach requires multiple observations of a teacher, which is not applicable in our setting.

Table 5
Mechanism: Teacher Gender or Other Teacher Characteristics?

	Academic Outcomes		Mental Stress					Social Acclimation and Satisfaction				
										Social Activities with Peers		
	Test Score (1)	Self-Assessment (2)	Depressed (3)	Blue (4)	Unhappy (5)	Pessimistic (6)	AES (7)	School Life Is Fulfilling (8)	Confident about Future (9)	Public Enrichment (10)	Private Recreation (11)	AES (12)
Female teacher × female student	.154***	.309***	−.179***	−.206***	−.150**	−.024	−.139***	.115**	.070	.021	.120**	.082**
	(.051)	(.050)	(.055)	(.063)	(.061)	(.051)	(.049)	(.054)	(.053)	(.048)	(.055)	(.036)
Subject fixed effects	Yes	Yes	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Female student × teacher control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Female teacher × teacher control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,202	18,601	6,672	6,657	6,667	6,648	6,620	6,688	6,722	6,586	6,561	6,466
R^2	.269	.088	.089	.086	.094	.070		.106	.156	.253	.250	

Note.—Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level. AES = average effect size; NA = not applicable.

** Significant at the 5% level.

*** Significant at the 1% level.

behavior, teachers may prefer students of a particular race, ethnicity, or gender, leading to different behaviors depending on student characteristics. For example, teachers may provide more feedback or time to students of the same gender (Jones and Dindia 2004).[15] Relatedly, female and male teachers may differ in their teaching styles or communication strategies, which may be better suited to students of their own gender.

Regarding student responses, it is possible that students may perceive teachers differently based on the teacher's gender. For example, the role model effect suggests that students feel more comfortable, inspired, and focused in class when a same-gender teacher is present, independent of the teacher's other qualities and behaviors.[16] Similarly, girls may feel less threatened by stereotypes if they have a female teacher. In this case, a same-gender teacher will be perceived by students as a counterargument to commonly held gender stereotypes—for example, that girls are less talented than boys in mathematics—that might otherwise hinder girls in their ability to perform at their maximum level.[17]

Our rich data allow us to directly test and separate these two classes of explanations. We first examine whether teachers behave differently toward female and male students. On questionnaires, students were asked to describe their interactions with their subject and head teachers on such dimensions as class questioning and provision of praise and criticism.[18] As shown in table 6, we find that compared with a male teacher, a female head teacher is less likely to criticize girls than boys, and a female subject teacher is more likely to question and praise girls than boys.

Next, we test students' perceptions of and responses to their teachers' gender. Table 7 presents evidence on the stereotype threat hypothesis and role model effects. Panel A reports the teacher gender effect on the students' belief that "boys are more talented in learning math than girls." Our results show that when mathematics is taught by a female teacher, female students are less likely to agree with the stereotype about a gender gap in learning math. Exposure to a female teacher who counters the stereotype appears

[15] In a similar view, Ferguson (1998) shows that in allocating class time and materials, teachers are more oriented to students of the same race.

[16] The literature on role model effects is relatively rich; see, e.g., Rothstein (1995), Canes and Rosen (1995), Jacobs (1996), and Diprete and Buchmann (2006).

[17] The stereotype threat hypothesis (Steele 1997; Spencer, Steele, and Quinn 1999) is supported by experimental evidence (Steele and Aronson 1995), in which female students underperform on a math test only when they are told that the test reveals gender differences.

[18] The four items ask students to rate how much they agree with the following statements on a scale from 1 (strongly disagree) to 4 (strongly agree): (1) the subject teacher always asks the student to answer questions in class, (2) the subject teacher always praises the student in class, (3) the head teacher always criticizes the student, and (4) the head teacher always praises the student. In the regression analyses, we normalize each variable to have a mean of 0 and a standard deviation of 1.

Table 6
Mechanism: Teacher's Behavior

	Subject Teacher's Class Questioning (1)	Subject Teacher's Praise (2)
A. Subject teacher's behavior:		
Female subject teacher × female student	.094***	.117***
	(.033)	(.035)
Female subject teacher	−.005	.002
	(.037)	(.034)
Female student	−.108***	−.065*
	(.038)	(.034)
Subject fixed effects	Yes	Yes
Block fixed effects	Yes	Yes
Student controls	Yes	Yes
Teacher controls	Yes	Yes
Observations	18,591	18,595
R^2	.160	.190
	Head Teacher's Praise	Head Teacher's Criticisms
B. Head teacher's behavior:		
Female head teacher × female student	.073	−.101*
	(.050)	(.051)
Female teacher	.100	.140*
	(.061)	(.073)
Female student	.007	−.215***
	(.041)	(.041)
Block fixed effects	Yes	Yes
Student controls	Yes	Yes
Teacher controls	Yes	Yes
Observations	6,820	6,825
R^2	.179	.074

NOTE.—Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level.
* Significant at the 10% level.
*** Significant at the 1% level.

to increase female students' confidence, which may contribute to our observed improvements in performance.

To examine the relevance of role model effects, we check whether a teacher's gender influences students' views about the usefulness of the subject for their future.[19] The role model effect implies that students feel more inspired when

[19] The survey item asks the students to rate, on a scale from 1 (strongly disagree) to 4 (strongly agree), how much the student agrees with the statement that "the subject is useful for my future."

Table 7
Mechanism: Student's Response

	Outcome: "Boys Are More Talented in Learning Math" (1)
A. Reduced gender stereotype threat:	
Female math teacher × female student	−.121*
	(.072)
Female math teacher	.060
	(.074)
Female student	−.152**
	(.064)
Block fixed effects	Yes
Student controls	Yes
Teacher controls	Yes
Observations	6,390
R^2	.070
	Outcome: "The Subject Is Useful for My Future"
B. Role model effect:	
Female subject teacher × female student	.240***
	(.040)
Female subject teacher	−.087**
	(.034)
Female student	−.025
	(.045)
Subject fixed effects	Yes
Block fixed effects	Yes
Student controls	Yes
Teacher controls	Yes
Observations	18,616
R^2	.109

NOTE.—Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

taught by a same-gender teacher. From table 7, panel B, we see that while male students' perceptions do not vary significantly by teacher gender, female students are more likely to consider a subject relevant and useful for their future when the subject is taught by a female. We interpret this result as evidence to support the idea that a female teacher serves as a role model and inspires her female students.

Taken together, we find evidence for both teacher- and student-based channels. Our findings reveal teachers' differential behaviors, depending on student gender, and their role in countering stereotypes and inspiring same-gender students. All results are robust to the inclusion of student characteristics and test scores. These findings, therefore, help to explain how female teachers improve girls' academic and noncognitive development.

C. Decomposing Teacher Gender Effects by Channel

To quantify how much each channel explains teacher gender effects—and their combined explanatory power—we use a decomposition method following Heckman, Pinto, and Savelyev (2013) and Gelbach (2016). Specifically, denote m^j_{icb} as the mechanism variable j and consider the channel estimation specification

$$\begin{aligned} m^j_{icb} = {} & \alpha^j_1 \mathrm{Femstud}_{icb} + \alpha^j_2 \mathrm{Femteach}_{cb} + \alpha^j_3 \mathrm{Femteach}_{cb} \\ & \times \mathrm{Femstud}_{icb} + \mathbf{W}'_{icb}\phi + D_b + \varepsilon^j_{icb}. \end{aligned} \tag{2}$$

Next, consider a long specification, that is, baseline equation (1) with all relevant mechanism variables:

$$\begin{aligned} Y_{icb} = {} & \beta'_1 \mathrm{Femstud}_{icb} + \beta'_2 \mathrm{Femteach}_{cb} + \beta'_3 \mathrm{Femteach}_{cb} \times \mathrm{Femstud}_{icb} \\ & + \sum_j \gamma^j m^j_{icb} + \mathbf{W}'_{icb}\phi + D_b + \varepsilon_{icb}. \end{aligned} \tag{3}$$

Gelbach (2016) shows that

$$\hat{\beta}_3 = \hat{\beta}'_3 + \sum_j \hat{\gamma}^j \hat{\alpha}^j_3, \tag{4}$$

which suggests that the mechanism j component of teacher gender effect is $\hat{\gamma}^j \hat{\alpha}^j_3$ and the remaining unexplained part is $\hat{\beta}'_3$. For each mechanism, we compute its explanatory power by $\hat{\gamma}^j \hat{\alpha}^j_3 / \hat{\beta}_3$.

Figure 1 presents our estimated decomposition of teacher gender effects on academic outcomes into teacher's questioning, praise, role model effect, reduced gender stereotype, and other factors. We find that for the effects of the subject teacher's gender, (1) the teacher's questioning behavior explains about 4.1% of the effect on test scores and 4.6% of the effect on self-assessment, (2) the teacher's praise explains about 6.5% of the effect on test scores and 4.1% of the effect on self-assessment, (3) the role model effect explains about 14.3% of the effect on test scores and 15.4% of the effect on self-assessment, and (4) the reduced gender stereotype threat explains about 2.3% of the effect on test scores and 3.0% of the effect on self-assessment. The four channels together can explain 27.2% of the effect of subject teacher's gender on students' test scores and 27.1% of the effect on students' self-assessment.

Figure 2 plots our estimated decomposition of teacher gender effects on noncognitive outcomes into the head teacher's praise and criticism behavior. We find that (1) the head teacher's praise explains about 5.4% of the effect on

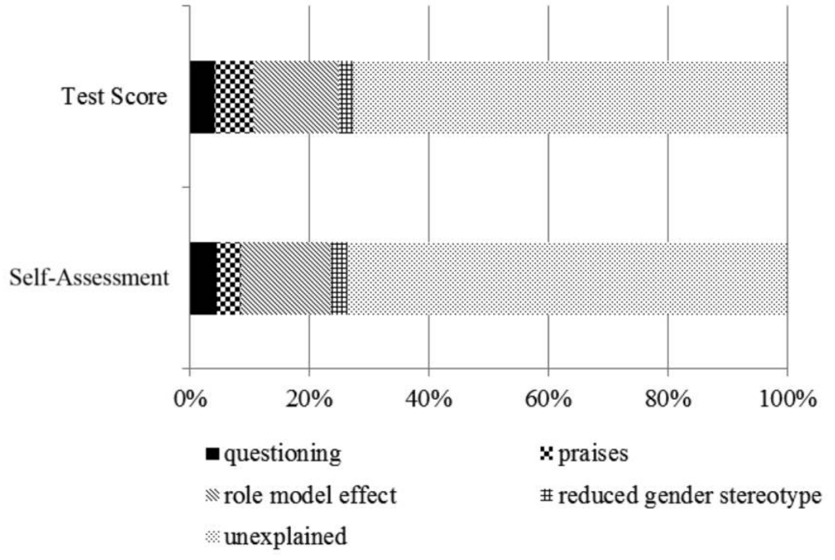

FIG. 1.—Decomposing teacher gender effects by channel: academic outcomes. This figure presents the estimated decomposition of teacher gender effects on academic outcomes. The effects on test score can be explained by teacher's questioning (4.1%), praise (6.5%), role model effect (14.3%), reduced gender stereotype (2.3%), and other factors. The effects on students' self-assessment can be explained by teacher's questioning (4.6%), praise (4.1%), role model effect (15.4%), reduced gender stereotype (3%), and other factors.

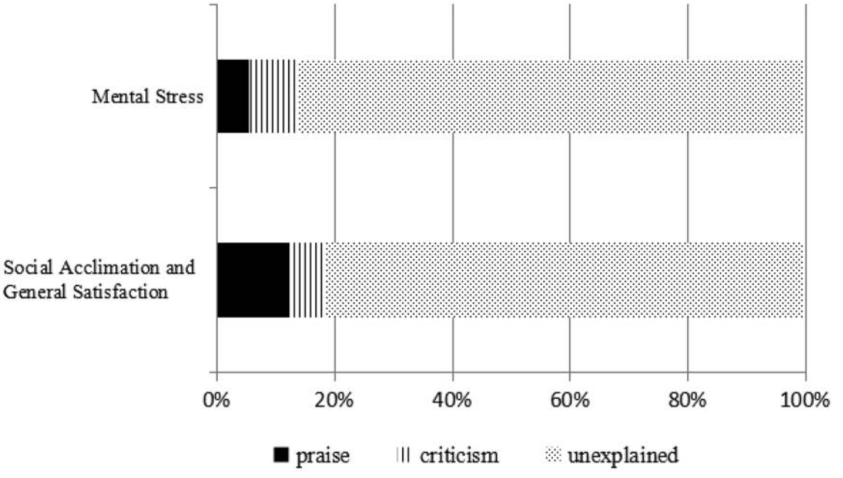

FIG. 2.—Decomposing teacher gender effects by channel: noncognitive outcomes. This figure presents the estimated decomposition of teacher gender effects on noncognitive outcomes. The effect on students' mental stress can be explained by the head teacher's praise (5.4%), criticism (8.4%), and other factors. The effect on social acclimation and satisfaction can be explained by the head teacher's praise (12.4%), criticism (6.2%), and other factors.

Table 8
Heterogeneous Effects on Test Score

	Mother's Education >9 Years (1)	Mother's Education ≤9 Years (2)	Ethnic Majority (3)	Ethnic Minority (4)	Living with Parents (5)	Parents Migrated Out (6)
Female subject teacher × female student	.102*	.266***	.160***	.458***	.150***	.349***
	(.061)	(.061)	(.047)	(.142)	(.051)	(.085)
Female subject teacher	−.044	−.080**	−.054**	−.154	−.045	−.119**
	(.037)	(.036)	(.026)	(.117)	(.031)	(.050)
Female student	.165***	.159**	.196***	.040	.190***	.046
	(.049)	(.066)	(.048)	(.091)	(.049)	(.099)
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,808	10,394	16,453	1,749	14,925	3,277
R^2	.315	.261	.277	.305	.273	.293

NOTE.—Test score is normalized by subject, grade, and school. Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

mental stress and 12.4% of the effect on social acclimation and general satisfaction and (2) the head teacher's criticism explains about 8.4% of the effect on mental stress and 6.2% of the effect on social acclimation and general satisfaction. Combined, the two proposed channels can explain 13.8% of the effect of the head teacher's gender on students' mental stress and 18.6% of the effect on students' social acclimation and general satisfaction.

VI. Heterogeneous Effects

Our main findings, which are reported in tables 3 and 4, capture the average effects of teacher gender on student outcomes. In this section, we examine whether the effects vary according to student background.[20] In the spirit of Cunha and Heckman (2007), students develop skills through various investments—for example, parenting, schooling, and environment—and the returns from these investments are interdependent. In this case, a teacher's influence may vary by the student's parental investment.

[20] We also examine the heterogeneous effects across teachers' teaching background (i.e., experience and education). We do not find that teaching experience matters for the teacher gender effect on gender differences in academic and noncognitive skills. For teacher education, we find that female teachers with a university degree and above have weaker effects on gender differences. Results are available on request.

Table 9
Heterogeneous Effects on Noncognitive Outcomes

	Depressed (1)	Blue (2)	Unhappy (3)	Pessimistic (4)	AES (5)	Depressed (6)	Blue (7)	Unhappy (8)	Pessimistic (9)	AES (10)
	Mother's Education >9 Years					Mother's Education ≤9 Years				
A. Mental stress:										
Female teacher × female student	−.151*	−.146	−.108	−.001	−.100	−.202***	−.206***	−.146**	−.037	−.150***
	(.090)	(.098)	(.107)	(.072)	(.077)	(.070)	(.068)	(.064)	(.065)	(.053)
Observations	2,788	2,788	2,788	2,788	2,788	3,832	3,832	3,832	3,832	3,832
R^2	.106	.108	.107	.105		.098	.097	.098	.074	
	Ethnic Majority					Ethnic Minority				
Female teacher × female student	−.166***	−.185***	−.123**	.017	−.114**	−.169	−.167	−.177	−.320**	−.208
	(.053)	(.060)	(.061)	(.051)	(.046)	(.190)	(.237)	(.177)	(.155)	(.148)
Observations	5,935	5,935	5,935	5,935	5,935	685	685	685	685	685
R^2	.086	.081	.092	.074		.192	.204	.163	.151	
	Living with Parents					Parents Migrated Out				
Female teacher × female student	−.171***	−.144**	−.106	−.013	−.109**	−.278*	−.349**	−.287**	−.148	−.264**
	(.060)	(.065)	(.065)	(.056)	(.051)	(.141)	(.142)	(.119)	(.132)	(.103)
Observations	5,401	5,401	5,401	5,401	5,401	1,219	1,219	1,219	1,219	1,219
R^2	.088	.088	.089	.074		.166	.174	.170	.156	

	School Life Is Fulfilling (1)	Confident about Future (2)	Social Activities with Peers: Public Enrichment (3)	Social Activities with Peers: Private Recreation (4)	AES (5)	School Life Is Fulfilling (6)	Confident about Future (7)	Social Activities with Peers: Public Enrichment (8)	Social Activities with Peers: Private Recreation (9)	AES (10)
	Mother's Education >9 Years					Mother's Education ≤9 Years				
B. Social acclimation and satisfaction:										
Female teacher × female student	.102	.034	−.056	.184*	.071	.029	.124*	.068	.130**	.090**
	(.067)	(.078)	(.088)	(.097)	(.059)	(.068)	(.071)	(.054)	(.052)	(.036)
Observations	2,713	2,713	2,713	2,713	2,713	3,753	3,753	3,753	3,753	3,753
R^2	.087	.152	.160	.150		.124	.164	.255	.220	
	Ethnic Majority					Ethnic Minority				
Female teacher × female student	.099*	.078	.022	.168***	.093**	−.129	.084	−.043	−.098	−.044
	(.057)	(.057)	(.046)	(.053)	(.036)	(.150)	(.186)	(.148)	(.141)	(.080)
Observations	5,804	5,804	5,804	5,804	5,804	662	662	662	662	662
R^2	.097	.158	.231	.238		.232	.249	.443	.364	
	Living with Parents					Parents Migrated Out				
Female teacher × female student	.068	.073	−.016	.170***	.075*	.080	.158	.108	.097	.110**
	(.061)	(.066)	(.049)	(.056)	(.040)	(.128)	(.123)	(.089)	(.107)	(.055)
Observations	5,281	5,281	5,281	5,281	5,281	1,185	1,185	1,185	1,185	1,185
R^2	.102	.164	.244	.247		.215	.215	.321	.291	
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE.—Student controls include student age, academic ranking in primary school, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family status, whether kindergarten was attended, whether a grade was repeated in primary school, and whether a grade was skipped in primary school. Teacher controls include age, years of schooling, experience, professional job title, and dummy variables indicating marital status, credential status, tenure status, whether the teacher graduated from a normal college, whether the teacher had prior experience in other schools, and whether a core subject is taught. Standard errors (in parentheses) are clustered at the block level. AES = average effect size.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

We explore differential effects across three dimensions: parents' education (proxied by mother's years of schooling), student ethnicity, and whether the student's parents migrated out.[21] As shown in table 8, we find that the female teacher effect is stronger for disadvantaged students. For example, we find that having a female teacher raises left-behind girls' test scores by 34.9% of a standard deviation relative to those of boys, compared with the increase of 15.0% of a standard deviation observed for girls with both parents around. We further find that coefficients on the female teacher × female student interaction are statistically different between the three pairs of subgroups: students whose mothers have more than 9 years of school versus those with less educated mothers, ethnic majority versus minority, and students who live with their parents versus those who were left behind by migrant parents.

In terms of noncognitive outcomes, we find that teacher gender effects (coefficients of the interaction term) are higher for disadvantaged students across all measures of mental health (table 9, panel A), but comparisons of coefficients between subgroups are not statistically significant. Differences between subgroups in social acclimation and general satisfaction are less pronounced (table 9, panel B). Because the effects do not differ statistically between subgroups, we should interpret heterogeneous effects with caution.

Overall, we find evidence that the female teacher effect on test scores is stronger for students from lower socioeconomic backgrounds, which may reflect less parental input. As a result, these students are more sensitive to their school environment and more influenced by their teachers in learning and development outcomes. These results reinforce the findings of other studies that disadvantaged students are more likely to be affected by the gender composition of their peers and class size (Angrist and Lavy 1999; Lavy and Schlosser 2011). An implication of the heterogeneous effect is that school-related policies that target disadvantaged students may have large returns.

VII. Conclusion

This paper sheds light on how teacher gender affects student outcomes. We use random assignment of students to waive concerns about self-selection in student-teacher matching. Our data allow us to extend student outcomes from traditionally focused academic achievements to noncognitive outcomes. A further advantage of our study is that various measures of teacher and student perceptions and behaviors are used to separate the possible mechanisms that drive the main effect of teacher gender.

Our findings suggest that having a female teacher improves both academic performance and noncognitive outcomes more for girls than for boys. On student test scores, our results show a reverse gender gap, which is enforced

[21] For mother's education, we use 9 years of schooling as the cutoff, since it is the sample median. Results are similar if we use father's education instead.

when taught by female teachers. This teacher gender effect is stronger among students from lower socioeconomic backgrounds. On noncognitive outcomes, girls tend to be more mentally stressed and less satisfied at school if they are taught by male teachers. However, female teachers can overturn this gender gap. In examining the mechanisms that drive these results, we find evidence of teacher behavior that varies depending on student gender as well as students' perception of teachers as role models and counterexamples to gender stereotypes.

Our study has a number of implications for educators and policy makers. First, our findings provide useful information for policy makers who seek to balance gender representation among middle school teachers. Second, our results provide insights for teachers regarding the differential impact of classroom actions, such as questioning and praising or criticizing girls versus boys.

Our findings can also be used to broaden our understanding of the nature and sources of boys' academic difficulties (e.g., Bertrand and Pan 2013; Cornwell, Mustard, and Van Parys 2013; Fortin, Oreopoulos, and Phipps 2015). It is commonly observed and recognized that boys are more likely to have behavioral and socioemotional difficulties, such as problems with self-control, that can limit their ability to obtain high grades. Our findings may help schools understand how teacher behaviors can contribute to a more gender-neutral school environment.

While this study offers insight into the role played by teacher gender at the middle school level, future research could investigate several important questions. For example, it would be valuable to study the long-term effects of teacher gender on students. Do these effects decline or persist to later education, and even adulthood? It would also be interesting to determine whether students are more affected by their most recent teachers or by teachers they had during pivotal stages of development, such as middle school. Finally, future studies could examine how teacher gender interacts with other school environment factors, such as peer influence. Answers to these questions would offer valuable insights into the effects of school context on children's development and long-term achievements.

References

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review* 96, no. 3: 847–62.

Angrist, Joshua D., and Victor Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on student achievement. *Quarterly Journal of Economics* 114, no. 2:533–75.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2015. The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics* 33, no. 1:63–89.

Bertrand, Marianne, and Jessica Pan. 2013. The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics* 5, no. 1:32–64.

Bettinger, Eric P., and Bridget Terry Long. 2005. Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review* 95, no. 2:152–57.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas Ter Weel. 2008. The economics and psychology of personality traits. *Journal of Human Resources* 43, no. 4:972–1059.

Campbell, Jay R. 2000. *Trends in academic progress: Three decades of student performance*. US Department of Education, Office of Educational Research and Improvement.

Canes, Brandice J., and Harvey S. Rosen. 1995. Following in her footsteps? Faculty gender composition and women's choices of college majors. *Industrial and Labor Relations Review* 48, no. 3:486–504.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics* 125, no. 3:1101–44.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys. 2013. Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources* 48, no. 1:236–64.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica* 74, no. 5:1191–230.

Cunha, Flavio, and James Heckman. 2007. The technology of skill formation. *American Economic Review* 97, no. 2:31–47.

Dee, Thomas S. 2004. Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics* 86, no. 1:195–210.

———. 2005. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review* 95, no. 2:158–65.

———. 2007. Teachers and the gender gaps in student achievement. *Journal of Human Resources* 42, no. 3:528–54.

Diprete, Thomas A., and Claudia Buchmann. 2006. Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography* 43, no. 1:1–24.

Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer. 1995. Do teachers' race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988. *Industrial and Labor Relations Review* 48, no. 3:547–61.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review* 104, no. 8:2567–91.

Ferguson, Ronald F. 1995. Can schools narrow the black-white test score gap? *The black-white test score gap*, ed. Christopher Jencks and Meredith Phillips, 318–74. Washington, DC: Brookings Institution.

Flossmann, Anton L., Rémi Piatek, and Laura Wichert. 2008. Going beyond returns to education: The role of noncognitive skills on wages in Germany. Technical report, University of Konstanz.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps. 2015. Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources* 50, no. 3:549–79.

Gelbach, Jonah B. 2016. When do covariates matter? And which ones, and how much? *Journal of Labor Economics* 34, no. 2:509–543.

Heckman, James J., and Tim Kautz. 2013. Fostering and measuring skills: Interventions that improve character and cognition. Technical report, National Bureau of Economic Research.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103, no. 6:1–35.

Heckman, James J., and Yona Rubinstein. 2001. The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review* 91, no. 2:145–49.

Hoffmann, Florian, and Philip Oreopoulos. 2009. A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources* 44, no. 2:479–94.

Holmlund, Helena, and Krister Sund. 2008. Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics* 15, no. 1: 37–53.

Jacobs, Jerry A. 1996. Gender inequality and higher education. *Annual Review of Sociology* 22:153–85.

Jencks. 1979. *Who gets ahead? The determinants of economic success in America*. New York: Basic.

Jones, Susanne M., and Kathryn Dindia. 2004. A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research* 74, no. 4:443–71.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. Experimental analysis of neighborhood effects. *Econometrica* 75, no. 1:83–119.

Lavy, Victor. 2010. Effects of free choice among public schools. *Review of Economic Studies* 77, no. 3:1164–91.

Lavy, Victor, and Edith Sand. 2015. The effect of social networks on students academic and non-cognitive behavioral outcomes: Evidence from conditional random assignment of friends in school. Working paper, University of Warwick and Hebrew University.

Lavy, Victor, and Analia Schlosser. 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics* 3, no. 2:1–33.

Neumark, David, and Rosella Gardecki. 1998. Women helping women? Role model and mentoring effects on female Ph.D. students in economics. *Journal of Human Resources* 33, no. 1:220–47.

Nixon, Lucia A., and Michael D. Robinson. 1999. The educational attainment of young women: Role model effects of female high school faculty. *Demography* 36, no. 2:185–94.

Rothstein, Donna S. 1995. Do female faculty influence female students' educational and labor market attainments? *Industrial and Labor Relations Review* 48, no. 3:515–30.

Segal, Carmit. 2013. Misbehavior, education, and labor market outcomes. *Journal of the European Economic Association* 11, no. 4:743–79.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. Stereotype threat and women's math performance. *Journal of Experimental Social Psychology* 35, no. 1:4–28.

Steele, Claude M. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52, no. 6:613–29.

Steele, Claude M., and Joshua Aronson. 1995. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69, no. 5:797–811.

Vincent-Lancrin, Stéphan. 2008. The reversal of gender inequalities in higher education: and on-going trend. In *Higher education to 2030*, vol. 1, *Demography*, 265–98. Paris: OEDC.

Winters, Marcus A., Robert C. Haight, Thomas T. Swaim, and Katarzyna A. Pickering. 2013. The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review* 34:69–75.