Performance Gains from Gender Match in Higher Education: Evidence from a Setting with Entrenched Gender Stereotypes*

Md Amzad Hossain†

Abstract

I use a novel and confidential administrative dataset from a highly rated economics program in Bangladesh to show that when matched with female teachers, female students perform better and are more likely to enroll in an economics master's program. Importantly, these gains come at no cost to male students. The quasi-random allocation of students to mandatory courses, where students cannot select either courses or instructors, alleviates the self-selection concerns. I rule out the explanation that the gain is driven by gender preference in teachers' assessment and show that female teachers' effectiveness in teaching female students is an important channel.

Keywords: Gender gap in STEM, higher education, teacher-student gender match, gender stereotypes

JEL codes: I23, J16, I24

*I am grateful to my supervisors at the University of Virginia, Sheetal Sekhri, Sandip Sukhtankar, and Leora Friedberg, for advising me and providing many helpful comments and discussions. I would also like to thank Sarah Turner, Eric Chyn, Gaurav Chiplunker, David Weil, Chao Fu, Jeffry Smith, Miguel Urquiola, Jonathan Colmer, Shan Aman-Rana, Marcos Rangel, Jonas Hjort, Jessica Goldberg, and seminar participants at the University of Virginia, University of Arkansas, Southern Economic Association Conference 2022, Washington Area Development Economics Symposium (WADES) 2022, and Eastern Economic Association Conference 2020, for insightful comments. I have no conflict of interest to declare. The University of Virginia Institutional Review Board (IRB-SBS) approved the protocol for this study (IRB protocol number: 3764).

†Assistant Professor, Department of Economics, University of Arkansas, Fayetteville, AR-72701, USA. Email: mdamzadh@uark.edu. Phone number: +1-434-466-1313

1 Introduction

Over the past three decades, female students in most developed countries have made remarkable gains in education relative to male students (Vincent-Lancrin, 2008; Goldin, Katz and Kuziemko, 2006). However, women's participation in science, technology, engineering, and mathematics (STEM) fields remains notably low. For instance, women earned only 36 percent of the total STEM degrees awarded and comprised only 28 percent of the STEM workforce in the U.S. in 2018.[1] The situation is similar for economics, which is often considered part of STEM.[2] Since 2000, the share of female Ph.D. students and female assistant professors in economics programs has remained stagnant at 35 percent and 30 percent, respectively (Lundberg and Stearns, 2019).

This begs the question of why so few women are in STEM fields. Aptitude and preparedness for college do not seem to explain this (Carrell, Page and West, 2010).[3] Social and environmental factors have been proposed as alternative explanations. One such factor is socio-cultural stereotypes (Ambady et al., 2001; Nguyen and Ryan, 2008); for example, many people believe that women are not as good as men in math and science. The literature on stereotypes documents that negative stereotypes can substantially impede performance (Aronson, Quinn and Spencer, 1998; Croizet and Claire, 1998; Spencer, Steele and Quinn, 1999).

A related second factor is the college environment (Hill, Corbett and St Rose, 2010), such as teachers or peers female students are exposed to in college (Carrell, Page and West, 2010; Hoffmann and Oreopoulos, 2009). For instance, teachers may behave differently toward students of different genders because of their beliefs about a student's innate abilities or relatability. In a related vein, Carlana (2019) finds that female students tend to underperform

[1]https://www.stemwomen.com/women-in-stem-usa-statistics

[2]In its list of science, technology, engineering, and math (STEM) degree program, the Department of Homeland Security (DHS) in 2012 included Quantitative Economics. https://www.ice.gov/sites/default/files/documents/Document/2014/stem-list.pdf

[3]The small differences that do exist in high school math and science can't explain this (Xie and Shauman, 2003), nor can the nearly nonexistent differences in college preparatory math and science courses (Goldin, Katz and Kuziemko, 2006).

in math and sort themselves into less demanding high schools when matched with teachers with stronger gender stereotypes. Similarly, students may respond differently to teachers of a different gender due to teaching styles, pedagogical techniques, or effective role modeling (Hoffmann and Oreopoulos, 2009). I focus on the role of teacher gender in this study.

In this paper, I estimate the gains from the gender matching between students and teachers at the college level in the context of a developing country. More specifically, I test whether female students gain academically when taught by female teachers. I focus on developing countries in which gender prejudices and stereotypes are an entrenched social phenomenon (Jayachandran, 2015). Societal norms and cultural values often discourage girls and women from studying STEM in many such countries, believing that it is a domain for boys and males. According to a 2015 UNESCO report, only 20 percent of STEM researchers in South and West Asia are women, whereas the corresponding worldwide figure is 30 percent.[4] The corresponding number for Bangladesh is 14 percent.

I focus on the academic performance of female college students for several reasons. First, economists have documented a positive correlation between undergraduate GPA and post-college earnings (Wise, 1975; Filer, 1983; Jones and Jackson, 1990), either through human capital or screening or both. Second, undergraduate GPAs are crucial in Ph.D. admissions decisions (Attiyeh and Attiyeh, 1997). Third, in the context of developing countries, in which information asymmetry is more prevalent (Luoto, McIntosh and Wydick, 2007), GPA often works as a crucial factor in job screening.[5] This is more true for jobs at universities and research institutes.[6] Finally, in deciding whether to write letters of recommendation for students, teachers use students' academic performance as one of the most important criteria. Hence, college exams are high-stakes tests.

Insofar as a higher academic achievement by female college students in STEM fields has the potential to increase participation in STEM higher education and STEM professions,

[4] http://unesdoc.unesco.org/images/0023/002351/235155E.pdf
[5] https://www.thedailystar.net/op-ed/politics/does-cgpa-matter-getting-job-1274815
[6] https://tbsnews.net/feature/pursuit/low-cgpa-it-end-job-opportunities-78283

my findings are of significant relevance. I uncover factors that improve the academic performance of female students in such settings. I estimate the effect of student-teacher gender matching using a novel administrative dataset from a highly selective economics program in Bangladesh. My data track every student by semester for all cohorts between 2007 and 2017. These data contain comprehensive information on students' course-level academic outcomes; medium-term academic outcomes, including the number of years repeated by students, the likelihood that they will complete their degree in four years, and their cumulative grade point average (CGPA); longer-term outcomes, such as the likelihood of enrolling in an economics master's program; and detailed demographic information for the corresponding instructors.

Two important identification challenges arise when investigating the role of a professor's gender in the academic outcomes of female students. First, students could choose their courses. Second, students could choose their professors. My setting allows me to address these selection pitfalls. Students in the program take mandatory courses each semester, for which they have no choice as to course or instructor, and I limit my analysis to these mandatory courses. I rely on two sources of variation for my identification: First, I use the within-cohort variation, whereby different teachers teach students of a particular cohort for different courses. Second, I exploit across-cohort variation in teacher gender over the same courses. I also control for student, course, cohort, teacher, and teacher-by-course fixed effects. This helps eliminate several concerns about omitted variable bias that arise in the literature on the effects of gender match on academic outcomes.

I find that female students perform relatively better in courses taught by female teachers. Female students' grades in the mandatory courses improve by approximately 9 percent of one standard deviation when the assigned course teacher is female. This effect is similar in magnitude to that of Carrell, Page and West (2010), and much higher in magnitude than that of Hoffmann and Oreopoulos (2009).[7] Male students benefit too from being taught

[7] Carrell, Page and West (2010) find that female students' performance in math and science courses improves by around 10 percent of a standard deviation when the course is taught by a female professor, whereas Hoffmann and Oreopoulos (2009) find an increase of at most 5 percent of its standard deviation in grade performance when matched with a same-sex teacher.

by female teachers, but the gains for female students are higher. While estimates of the gain from pairing female students with female teachers are significant for math-intensive courses such as mathematics, statistics, and econometrics, the increase in test scores is twice as high for other courses than for those in mathematics, statistics and econometrics.[8]

I also estimate the effect of the proportion of mandatory courses a student takes from female professors on degree completion time and the number of years a student repeats. I find that the proportion of mandatory courses taken by a female student and taught by female professors increases the likelihood of graduating with a regular cohort for female students and results in a higher GPA. In addition, more exposure to female teachers during the bachelor's degree increases the likelihood of enrolling in an economics master's program for female students. I bolster my findings by performing a number of robustness checks to rule out the possibility that the assignment of teachers is based on the gender composition of the class or the performance of the female students.

I then explore the mechanisms through which female students benefit from taking classes with female teachers. A potential mechanism is that the improvement in female students' performance may be due to bias in teachers' assessment of tests (Bar and Zussman, 2012; Hanna and Linden, 2012; Burgess and Greaves, 2013). A feature of this study's setting, which allows me to address this, is that the students take a non-blindly graded midterm and a blindly graded final exam for each course. I do not find any evidence that teachers favor students of the same gender.

Once I establish that the gain from matching is not driven by gender bias in teacher's assessment, I look at alternative mechanisms. Are female teachers more effective in teaching female students? Or do female teachers serve as "role models" for their students? Earlier studies mostly cite the role-model effect as the driving mechanism without going into much detail (Carrell, Page and West, 2010; Hoffmann and Oreopoulos, 2009; Bettinger and Long, 2005). While it is very much difficult to distinguish between the role-model effect

[8]These other courses include, among others, introductory microeconomics and macroeconomics, intermediate microeconomics and macroeconomics, international trade, development economics, etc.

and the teacher-effectiveness channel, I provide suggestive evidence that female teachers' effectiveness in teaching female students is an important channel through which same-sex teacher assignments improve female students' academic achievements. I show this in two ways. First, I produce several pieces of evidence that the role-model effect cannot explain; Rather, the teacher-effectiveness channel can more persuasively explain those. Second, I use the insights from the theoretical models in the literature that formalizes the notion of "role-model effects" as distinct from teacher effectiveness. Using the testable implications from these models (Gershenson et al., 2018), I show evidence that teacher effectiveness is an important channel.

I make a novel contribution to the literature on student-teacher gender matches in three ways. First, I provide the first causal estimates of gender matching between students and teachers at the college level in the context of a developing country. Several studies investigate the role of teacher gender at the post-secondary level in developed-country settings (Rothstein, 1995; Neumark and Gardecki, 1998; Bettinger and Long, 2005; Hoffmann and Oreopoulos, 2009; Carrell, Page and West, 2010), but to the best of my knowledge, no such study exists for developing countries. Because stereotypes and discrimination are much more pervasive in developing countries (Jayachandran, 2015), therefore, my paper fills an important gap. Second, by focusing on college, I examine to what extent the gender-interaction effects observed in schools in developing countries (Muralidharan and Sheth, 2016) are present at later ages. This is important since academic performance at the college level has more immediate implications for STEM higher education and, consequently, the STEM labor market. Finally, the quasi-random allocation of students in my setting, whereby the students cannot choose courses or instructors, allows me to estimate the effect of student-teacher interaction on different student outcomes while avoiding the self-selection problem often encountered in prior research (Carrell, Page and West, 2010).

I also contribute to the literature on gender bias in teacher assessments. Several studies have exploited the difference in the test scores of blind and non-blind exams to document evidence of such biases in teachers' grading, a technique pioneered by Lavy (2008). Several studies find substantial bias against boys in teacher assessments (Lavy, 2008; Falch and Naper, 2013; Cornwell, Mustard and Van Parys, 2013; Lindahl, 2016), while others document no such gender bias (Hinnerich, Höglin and Johannesson, 2011). In this paper, I go one step further and, following Lindahl (2016), ask whether teachers favor students of their own gender. However, unlike Cornwell, Mustard and Van Parys (2013), who find that female teachers are less generous when grading the non-blind exams of female students, I do not find evidence to support the view that teachers either favor or disfavor students of the same gender. My paper also differs from earlier studies, in which blind and non-blind exams are graded by different graders. In my setting, both exams are graded by the same person.

My paper also adds value to the literature by investigating the mechanism through which female students benefit from being paired with female teachers. Previous researchers cite the role-model effect as the key mechanism without going into much detail (Carrell, Page and West, 2010; Hoffmann and Oreopoulos, 2009). However, in this paper, I provide suggestive evidence that female teachers' effectiveness in teaching female students can also be an important channel.

Lastly, my paper makes a direct contribution to the debate concerning increasing female representation in a male-dominated field such as economics. Increasing female representation in STEM fields has frequently been advocated (Hill, Corbett and St Rose, 2010). This paper studies the impact of matching female students with female teachers, and whether such matching can influence student academic performance. Against the backdrop of ongoing debates and legislative changes with respect to affirmative action in developing countries, this topic warrant investigation.

The remainder of the paper is organized as follows. Section 2 describes the dataset and Section 3 discusses the empirical strategy employed in this paper. In Section 4, I present my main results. Section 5 investigates the mechanisms, and Section 6 concludes.

2 Background and Data

My analysis is based on the administrative data from the economics department of one of the most reputed public universities in Bangladesh. Admission into this university is highly competitive, and its economics program is equally selective. Fewer than one out of 200 applicants are accepted for admission into the department, for a total of 150 seats. The department offers a four-year degree in economics, where each student has to complete eight semesters of coursework. Each student is therefore expected to graduate in four years, although some students take more time due to repetition.[9] Each semester, a student takes a combination of mandatory and optional courses. The number of mandatory and optional courses is the same for all students of the same cohort, although it may differ across cohorts. A student failing to achieve the minimum threshold GPA at the end of her year must repeat the year with the subsequent cohort. Each student in the same cohort has the same course instructor for a particular course, and multiple sections of the same course taught by different professors are not offered. Thus, for a given course, the student cannot select the course teacher.

Each course is graded on a scale of 100 points. This is composed of two parts – Sessional exams and Final exams – each worth 50 points. The sessional exams consist of one midterm examination (30 points) and in-class exams (20 points). The Final exam carries 50 points and is held at the end of the semester. The same instructor grades both the sessional and final exams. But, there is an important distinction between these two types of exams. The sessional exams are non-blindly graded exams, where the instructor – who is also the grader, knows the identity of the students. The final exams, on the other hand, are blindly graded exams where the instructor-cum-grader is unaware of the student's name and gender.[10]

[9]This system is different from a four-year college degree in countries like the U.S., where a student must choose his/her major in the sophomore (second) or junior (third) year. In Bangladesh, students must choose their major before enrollment into colleges, and once enrolled, they cannot switch to other disciplines.

[10]For the first four cohorts, the exams were sent to a second examiner for additional blind grading. In that case, the final grade was the average of the two grades.

I collect detailed information on course outcomes that include scores received in both the sessional exam and the final exam for every course offered to every student enrolled in the department from January 2007 to January 2017. My sample consists of approximately 1500 students from 11 cohorts. Approximately 35 percent of the students in my sample are female (Table I). My data also includes information on the academic records of students repeating a year. Further, I collect students' baseline characteristics at entry such as merit score in the entrance exam, the major in high school, and whether the student was admitted to reserved seats.[11]

Academic performance measures in the data consist of Sessional grades (50 points), Final grades (50 points), total grades (100 points), and grade points (on a scale of 4.00) for every individual student by course and semester. Students at the Department of Economics are required to take a mix of approximately 35 mandatory and optional courses, but in my analysis, I use the data for the mandatory courses as there is no selection possibility in these courses.

Grades points are assigned on an A+, A, A, B+, B, B-, C+, C, D, F scale, where an A+ is worth 4-grade points, an A is 3.75-grade points, an A is 3.50-grade points, a B+ is 3.25-grade points, and so on (see Appendix Table A.1 for a detailed description). The sample cumulative grade point average (CGPA) for female and male students is 3.14 and 2.95, respectively. I standardize both the course grades and grade points, such that each variable has a mean of zero and a variance of one, within each course offered to a particular cohort.

For each of the courses taught during different semesters in the department, individual instructor-level data were obtained from the department. I collect data on each teacher's gender, academic rank, highest level of education (M.A. or Ph.D.), and years of tenure in the department. These were merged with the data on students' academic achievement. During the period I study, 66 different teachers taught all courses. Of these, 30 percent

[11]Five percent of the total seats are reserved for the students belonging to an ethnic minority or to freedom fighter families and to the children of the university staff.

(20 of 66) were female and taught nearly 42 percent (101 of 243) of the mandatory courses. In this sample, 50 percent of the female faculty members had a Ph.D. degree, while the corresponding number for the male faculty members is approximately 43 percent.

3 Empirical Strategy

3.1 Selection into Courses and Professor Choice

The existing literature faces two main identification challenges in estimating the impact of female students' exposure to female teachers. First, a student could choose their courses. As explained earlier, in my setting, students must take certain mandatory classes each semester, in which they have no choice. My primary analysis focuses on these classes. Another advantage of restricting my analysis to the mandatory courses is that the class size of those courses is large—typically more than 100 students—which reduces the likelihood that my results are based on anomalous circumstances (Hoffmann and Oreopoulos, 2009).

The second selection issue is that students could choose their instructor if multiple instructors were to teach the same course. Again, this problem does not arise in my setting. This is because every student in the same cohort is exposed to the same teacher for a particular course. This instructor is assigned by the department on the basis of expertise and experience. Therefore, the identification of female student-teacher interactions comes from the rich cross-cohort variation in teachers' gender for different courses (see Figure 1 for a stylized illustration of this). The cross-cohort variation in teachers' assignments to different courses is due to several factors, which I treat as exogenous. First, there is attrition due to retirement or personal reasons. For example, a large number of young teachers leave the department to pursue graduate studies. Second, when several teachers specialize in a particular field, they are assigned relevant courses to teach in alternate years. Third, sabbaticals of different natures also affect which professors teach mandatory courses in a particular semester.

The fact that I limit my analysis to mandatory courses, coupled with the fact that students have no role in selecting teachers, gives me a quasi-experimental design to estimate causal effects. I also provide empirical support for my claim that the allocation of students to classrooms is random. To this end, I compare the female students taught by male and female teachers and the male students taught by male and female teachers. If the assignment of the students to male and female teachers were random (not based on observable student characteristics), we would expect the groups of male students and female students to be similar in terms of observable characteristics when they are taught by male and female teachers, respectively. I report the sample mean for all students when they are taught by male and female teachers in Table II. I show the characteristics of female students by the gender of the teacher in Panel A and the characteristics of male students by the gender of the teacher in Panel B. We do not observe any significant difference in the mean characteristics of the two groups of students by teacher's gender. This suggests that students are not more likely to be matched with a same-sex teacher on the basis of observable characteristics.

I do several other robustness checks to show student assignments to teachers are random. For instance, a concern may be that female teachers could be assigned to specific classes based on the gender composition of the class or the performance of the female students. In order to rule this out, I run regressions to show that the proportion of mandatory courses taught by female teachers in a single semester is not influenced by the gender composition of the class (Appendix Table A.2) or female students' performance in the preceding semester (Appendix Table A.3).

3.2 Empirical Specification in the Quasi-Experimental Setting

Every student belonging to a particular cohort is assigned only one teacher for a specific mandatory course. However, across cohorts, this same course may be taught by a female or a male professor. Thus, I rely on two sources of variation: I use the within-cohort cross-sectional variation, where students from the same cohort are taught by different teachers for different courses. In addition, I use the across-cohort variation, in which students from different cohorts are taught by different teachers for the same course. To clarify my identification strategy, consider the following: Students of the first cohort enter the program. They are obligated to take a certain number of mandatory courses each semester, from first to eighth. The order of these courses is fixed by the department, and students cannot take those courses in a different order. Students of this cohort take mandatory courses, some of which are taught by male teachers and some by female teachers. This is the cross-sectional variation. Now, students of the second cohort enter. They are offered coursework in a similar order. The only thing that is changed is that different teachers are assigned for their courses relative to the first cohort (due to the reasons outlined before). This gives rise to circumstances where a course that was taught by a male teacher for the first cohort is taught by a female teacher for the second cohort. This across-cohort temporal variation in the gender of instructors teaching the same courses allows me to account for student, teacher, course, and cohort fixed effects and teacher-by-course fixed effects.

The rich set of fixed effects helps in identification in the following way. First, there is a possibility that students taking courses taught by female teachers are systematically different from those who are not, regardless of their gender. While not likely in my setting, this absolute sorting is accounted for by including student fixed effects in my regression model. Second, it is possible that female students take courses from teachers who are systematically different from other teachers. I eliminate this concern by including individual teacher fixed effects. Third, it might be the case that male and female teachers are systematically different in teaching certain courses. To mitigate this concern, I include teacher-specific course fixed effects. Finally, students within a particular cohort-semester are taking the same courses and are exposed to similar classroom-specific shocks, such as the time of day or different other external disruptions. I include classroom (cohort-semester-course) fixed effects to control for this concern. Absorbing these fixed effects helps me compare the outcomes of male and female students of the same cohort when matched with female teachers compared to those taught by a male teacher.

To estimate the effect of student-teacher interaction, I start with the regression model of the following form:

$$y_{ijkst} = \alpha_0 + \alpha_1 \times FT_{jkst} + \alpha_2 \times FS_{ikst} + \alpha_3 \times FT_{jkst} \times FS_{ikst} + X_{ijkst}\beta + u_{ijkst} \quad (1)$$

where I index students by i, teachers by j, course by k, cohort by s, and semester by t. FT_j and FS_i are dummy variables taking values equal to one if teacher j and student i are female. X_{ijkst} is a set of observable variables, such as the teacher's experience and the highest level of education. Finally, u_{ijkst} is the set of unobservables. I am primarily interested in α_3, which shows the differential effect between female and male students after being exposed to a female teacher for a particular course. A positive value of this parameter implies that female students perform relatively better than male students when taught by a female teacher.

Including all these fixed effects in equation (1), I obtain a regression model of the following form:

$$y_{ijkst} = \alpha_3 \times FT_{jkst} \times FS_{ikst} + \gamma_i + \lambda_j + \phi_{kst} + u_{ijkst} \quad (2)$$

where γ_i are the student fixed effects, λ_j are the teacher fixed effects, and ϕ_{kst} are cohort-semester-course fixed effects (henceforth, classroom fixed effects). Including these fixed effects allows me to overcome many threats to internal validity. However, when we include these fixed effects in Equation (2), student and teacher variables at the level form are dropped, due to multicollinearity with either of the fixed effects. The parameter α_3 estimates the effect of student-teacher interaction.

One remaining possibility is that female students gain from being exposed to female teachers because female teachers favor female students in grading. To test for gender bias in teachers' assessments, I exploit a rule that dictates that while the sessional exam (mid-term) of a course is non-blind, the final exam must be blind. Therefore, it may safely be assumed that scores in the blind test would not be influenced by any bias resulting from graders' gender stereotypes. On the contrary, scores in the non-blind exam may be affected by biases resulting from graders' stereotyping and discrimination. Since both the blind and non-blind exams test the same cognitive skills, I can use the blind score as a valid counterfactual of the non-blind score, which might be influenced by stereotyped discrimination.[12]

The sessional exams and the final exams are different in some respects. First, the sessional exams do not cover the complete syllabus of the course, whereas the final exams are based on the full syllabus. Second, the sessional exams are shorter (midterms take two hours, whereas the final exams take three hours). Yet, at the same time, they bear many similarities. First, both the sessional (non-blind) and final exams (blind) measure students' academic performance. Second, both exams in a particular course are set by the same professor, and although the final questions are considerably longer, the questions' nature is quite similar. Third, both exams take place in a similar type of exam-taking environment. This reduces the concern that differences in the exam environment that might be correlated with certain stereotyped feminine characteristics (e.g., possible higher anxiety levels) pose threats to the identification. All these similarities imply that both blind and non-blind tests measure the same cognitive skills. Therefore, once I control for the exam (i.e., blind versus non-blind) fixed effects, blind scores could be construed as a valid counterfactual of the non-blind score.

I explore the difference between male and female students' gaps between the scores of the blind and non-blind exams in order to capture potential gender bias. To control for systemic differences between the two tests, I control for examination-type fixed effects that control for any time-invariant unobservable characteristics. More precisely, to examine

[12] In this regard, my identification strategy resembles that of Lavy (2008), which was in turn inspired by Goldin and Rouse (2000) and Blank (1991).

whether teachers favor students of their gender, I run the following model,

$$y_{ijkste} = \alpha_3 \times FT_{jkst} \times FS_{ikst} \ + \ \alpha_4 \times A_e \ + \ \alpha_5 \times FT_{jkst} \times A_e \ + \ \alpha_6 \times FS_{ikst} \times A_e \\ + \ \alpha_7 \times FT_{jkst} \times FS_{ikst} \times A_e \ + \gamma_i + \lambda_j + \phi_{kst} + \ u_{iec} \tag{3}$$

where A_e is an indicator variable taking a value of 1 if the exam is blind (anonymous) and 0 otherwise (non-blind). The main parameter of interest is α_7, which shows the difference between the blind (anonymous) scores of male students and those of female students when they are matched with a female instructor, given the respective difference in the non-blind scores. A positive value of α_7 indicates that female teachers favor female students while grading.[13]

Finally, I measure how instructors' gender influences longer-term outcomes, such as the number of years repeated, the likelihood to graduate with the regular cohort or a student's cumulative grade point average. To this end, I run the following model, which is a variant of the regression model (2)

$$y_{ist} = \lambda_s + \alpha_1 \times FS_{is} + (\alpha_2 + \alpha_3 FS_{is}) \frac{\sum FT_{jst}}{n_{ist}} + u_{ist} \tag{4}$$

where y_{is} is the outcome of student i belonging to cohort s upto semester t. The outcome variable is an indicator variable taking values of 1 if a student graduates with their regular cohort or is a count variable indicating the number of years repeated by the student. $\sum FT_{jst}/n_{ist}$ is the share of mandatory course teachers j who were female for student i belonging to cohort s up to semester t. I include this variable in the model to help estimate the average impact of being exposed to more female teachers in mandatory courses. As in Equation (1), I am primarily interested in α_3, which estimates the differential effect across male and female students of being assigned to more female teachers in the mandatory courses.

[13] Though unlikely, it is possible that teachers speculate about students' genders from their handwriting in the test. In order to address this, I use yet another feature of the testing system, which I discuss in a later section concerning robustness tests.

4 Main Results and Discussion

4.1 Grade Performance

Estimates of the effect of the interaction between female students and female teachers on standardized test scores are reported in Table III. In Column 1, I report a basic specification without any fixed effects. In Column 2, I control for baseline characteristics of the students, such as merit scores in the entrance exam, fields of study in high school, and whether the student was admitted as part of a quota. I add an array of fixed effects in the subsequent columns to identify the parameter of interest: Column 3 includes student fixed effects, and Column 4 adds teacher fixed effects. I add classroom (cohort-semester-course) fixed effects in Column 5 and course by teachers' gender fixed effects in Column 6. I cluster the standard errors at the classroom (cohort-semester-course) level.

Table III shows that the estimate of the female student-teacher interaction is statistically significant across all specifications. In specifications that include student fixed effects, female students' grades in mandatory courses increase by about nine percent of a standard deviation when a female instructor teaches the course. Once student fixed effects are included, these results are robust to the inclusion of teacher, cohort-semester, or course fixed effects. This increase in grade performance is much higher compared to that estimated in Hoffmann and Oreopoulos (2009), who found that a same-gender teacher improves the average grades of female students by, at most, five percent of a standard deviation.

Two other results are worth noting, although the identification is not as strong as when we include the fixed effects. First, in column 1, where we do not include student fixed effects, we find that female students perform better than their male counterparts regardless of the gender of the teachers. This result reflects the finding of Table I, which shows that female students perform much better in terms of grade performance and medium-term outcomes. Second, male students also perform better when they are matched with female teachers. Male students' grades increase by 2.8 percentage points of a standard deviation when they take classes with female teachers instead of male teachers. The estimates of the gain are much higher when we include student fixed effects (Column 2 of Table III).

I then examine whether the increase in female grade performance from being matched with a female teacher differs across types of subjects. To this end, I divide all the mandatory subjects into two groups: (a) maths, statistics, and economics, and (b) all other courses. The results are documented in Table IV, which demonstrates that although the estimates of the gain from pairing female students with female teachers are significant for both groups of subjects, the increase in test scores is twice as high for other courses than those in mathematics, statistics, and econometrics. This result is qualitatively similar to Hoffmann and Oreopoulos (2009), who found that gains from the interaction tend to be higher in social science courses compared to math courses.

I also shed light on which groups of students benefit more from gender-matching, in terms of ability. To this end, I run a quantile regression model of student-teacher interaction. The results are depicted in Figure 2. It is evident that female students belonging to the middle tier of the distribution of normalized scores gain the most when they are exposed to female teachers. I supplement this distributional analysis by estimating the interaction effect at different quantiles of the merit score in the entrance exam and find consistent results (Appendix Table A.5).

4.2 Other Medium- and Longer-Term Outcomes

I also look at the impact of matching female students with female teachers on medium-term outcomes such as cumulative grade point average, the likelihood of being retained in the regular cohort, and the number of years repeated by the students. I run the model specified in Equation 4 for these outcomes. As seen in Table V, the estimates of the interaction of female students and the proportion of female teachers in all mandatory courses on cumulative grade point average are statistically significant. Similarly, as students are exposed to more and more female teachers as they progress through college, the probability of being retained in their regular cohort increases, and the number of years repeated declines. For instance, a one percentage point increase in the share of female teachers is associated with a 0.4 percentage point increase in the relative likelihood of passing with the regular cohort for female students compared to their male counterparts.

I then look at longer-term outcomes such as the likelihood of enrolling into an economics master's program and cumulative grade point average in master's conditional upon enrollment. Column 4 of Table V shows that a one percentage point increase in the share of female teachers in the compulsory undergraduate classes increases the relative likelihood of female students enrolling in an economics master's program by 1.14 percentage point, compared to male students. Conditional on enrollment, however, we do not see any significant effect of female student-teacher matching on female students' academic performance in the master's degree.

One point to note is that female students' exposure to female teachers in the first year alone has no impact on academic achievement in the later years. Table XII shows that the coefficient of the interaction of female students and the proportion of female teachers in first-year compulsory courses is statistically insignificant for most of the medium and longer-term outcomes such as Cumulative grade point average (CGPA), or the number of years repeated by a student, or the probability of enrolling into an economics master's program.[14] These results are similar to those of Carrell, Page and West (2010), who shows that having a high proportion of female teachers in the introductory classes does not affect longer-term outcomes such as withdrawal from the program within the first two years of enrollment or likelihood of graduating with a STEM degree. On the contrary, as Table V suggests, it is the continual interaction of female teachers and female students that affect medium-term outcomes such as overall CGPA, or the likelihood of graduating with the regular cohort, and longer-term outcome such as the likelihood of enrolling into economics master's program. These results suggest that it is the continuous exposure of female students to female teachers, rather than

[14]The only exception is the probability of passing with the regular cohort, for which more exposure to female teachers in the first-year compulsory courses actually worsens the outcome significantly.

the one-shot exposure, that affects female students' medium and longer-term gains.

However, one should interpret the results of the medium-and longer-term analysis with one caveat in mind: the identification is not as strong as was in the case of my main analysis in Section 4.1 since I could not control for student and teacher fixed effects. Regardless, this analysis suggests longer-term gains from the same-sex student-teacher pairing. One result that particularly stands out is that having more female teachers leads to an increase in the likelihood of enrolling in a master's program in economics for female students. Female under-representation in graduate economics courses has been subject to much attention in recent times (Lundberg and Stearns, 2019). My results suggest that hiring more female teachers can reduce the gender gap in economics higher education without harming the boys.

5 Possible Mechanisms

In this section, I explore the mechanisms through which female students benefit from taking classes with female teachers. First, I test whether the gains from matching are due to bias in teachers' assessment of test scores. Then, I examine whether the gains from matching can be explained by student and teacher observable characteristics. Finally, I discuss suggestive evidence about two potential mechanisms: the role-model effect and differential teacher effectiveness.

5.1 Is There Grading Bias?

Several papers document bias in teachers' assessments and grades.[15] I, therefore, aim to test whether the improved grade performance of female students, when matched with female teachers, is driven by assessment bias. Following the seminal paper of Lavy (2008), I use the scores obtained by students on their blind and non-blind exams to test whether such biases in teachers' assessments exist. Table VI shows the effect of the match between female students

[15]See, for example, Bar and Zussman (2012); Hanna and Linden (2012); and Burgess and Greaves (2013).

and female teachers on normalized test scores for blind and non-blind exams separately. It is evident that the interaction effect of female students taught by female teachers is quite similar for both blind and non-blind exams. Female students' performance in mandatory courses increases by approximately nine percent of a standard deviation in the non-blind exam when a female is an instructor. In contrast, the corresponding increase is nearly seven percent of a standard deviation in the blind exams. The Chow test suggests that this difference is not significant.

To test whether estimates of the interaction between female students and female teachers are driven by grading bias, I use a different approach by estimating a triple-difference model by interacting the blind exam dummy with a female student dummy and female teacher dummy (Equation 3). The result is reported in the first column of Table VII. The main parameter of interest is the coefficient of the triple interaction of female students, female teachers, and blind exam dummies. It turns out that the estimate of the triple interaction is insignificant, which implies that female teachers do not favor students of their gender when grading.

However, it is important to note several caveats regarding my analysis of blind and non-blind exams. First, the non-blind exams take place in the middle of the semester, while the blind exams are given at the end. This may bias my estimates if male and female students have different learning trajectories after the non-blind exams. I attempt to control for this by including student-by-exam-type fixed effects. Second, although the blind exams are anonymous, since the same examiner grades both the blind and non-blind exams, it is possible that the grader can infer students' identities from their handwriting. However, I believe this does not pose a threat to my study for several reasons. One, with a class size of more than 100, it is difficult to remember individual handwriting. Two, if the practice of inferring identity from handwriting is teacher-specific, that would be absorbed in teacher fixed effects. Three, for the first four cohorts in my study, the blind exams were graded by two examiners—the course teacher and an external examiner. The final grade was the average of the two grades. If the course teacher favored a student by inferring identity from handwriting, the interaction effect of the match between female students and female teachers would be different in the double-blind and single-blind exams for these cohorts. Yet, as can be seen from Appendix Table A.4, the estimate of the interaction is statistically similar in both single- and double-blind exams, and the Chow test of the equality of the two coefficients cannot be rejected. These facts suggest that inferring gender identity from a student's handwriting should not be a serious threat in the context of my study.

5.2 Do Students' and Teachers' Observable Characteristic Matter?

In this section, I examine heterogeneity in the treatment effect by teachers' and students' observable characteristics, with the goal of identifying mechanisms further. Table XI indicates that the gains from matching vary significantly with teachers' experience and rank. For instance, the gains are much higher when female students are matched with female teachers with higher hierarchical ranks (Columns 1-3). Notably, the interaction effect is negative when entry-level teachers teach courses and are successively higher for assistant or associate professors and full professors. Similarly, I find that the gains from gender matching are higher in courses taught by teachers with more experience (Columns 4-6). Finally, we see that estimates of the interaction of female students and female teachers are approximately 21 percent higher when the teacher has a Ph.D. than when the teacher does not (Column (8), Table XI). However, the Chow test reveals that the difference is not statistically significant. In Table VIII, I show these results by including the full set of interactions, in which I interact female student dummy, female teacher dummy, and teachers' observable characteristics. It is evident that gains from matching are driven by the rank and experience of the teachers.

I also examine whether observable student characteristics drive the gains from gender matching. To this end, I interact several student-level variables with the teacher gender dummy and student gender dummy to see whether the treatment effect of gender matching varies by student characteristics such as merit score and merit position in the entry exam, high school major, and quota status. However, Table IX shows that none of the estimates of the triple interaction variable are statistically significant. This suggests that student observable characteristics are not driving the female student-teacher interaction effect.

5.3 Role Model Effect VS Teacher Effectiveness

My results suggest that the improvement in female students' academic achievements is not driven by assessment bias. Previous literature offers two other classes of explanations (Dee, 2004). The first can be generalized as a role-model effect (Hoffmann and Oreopoulos, 2009). According to the role-model effect, an instructor's gender identity, by its very presence creates a role model that encourages students to exert more effort (King, 1993). The second class of explanations is that female teachers may be more effective in teaching female students. For example, female teachers may alter their pedagogical styles to meet the needs of female students, rendering female students more receptive to the material being taught.

Though it is difficult to disentangle the role-model effect from teacher effectiveness, previous literature identifies the role-model effect as the main mechanism through which gender interaction benefits female students (Carrell, Page and West, 2010; Hoffmann and Oreopoulos, 2009; Bettinger and Long, 2005). However, in this paper, I provide suggestive evidence that teacher effectiveness is also an important channel. I show this in two ways. First, I highlight some findings the role model effect cannot explain; the teacher-effectiveness channel seems to more plausibly explain those findings. Second, I borrow insights from the theoretical models in the literature, which differentiate between the role-model effect and teacher effectiveness in the case of same-race pairing Gershenson et al. (2018). Using their testable predictions, I find suggestive evidence that teacher effectiveness is an important channel through which same-sex teacher matching improves female students' academic outcomes.

5.3.1 Findings The Role-model Effect Cannot Explain

(a) Gains from matching are negative for entry-level female teachers

As discussed above, Table XI shows that the interaction effect is negative when entry-level teachers teach courses and is successively higher for assistant or associate professors and full professors. While one can argue that senior/experienced female teachers are better role models (Breda et al., 2020), the role-model effect cannot explain why female students perform worse when they are paired with entry-level female teachers. It seems unlikely that entry-level/young female teachers create a negative role model for female students, such that students exert less effort. If anything, female students might be encouraged more by a younger female teacher in a developing country, since they see a female who is marginally older overcoming the gender norms and stereotypes to become a teacher in an elite university. A more plausible explanation might be that female teachers with more teaching experience have learned over time how to bring the best out of female students and incorporate this in their pedagogical style (Ladson-Billings, 1995; Foster, 1990).

(b) Gains from matching are higher for below-median students

As seen earlier in Figure 2, the gain from female teacher-student matching is higher for students in the middle tier of the normalized score distribution; the interaction effect is actually negative for students who are above the 80th percentile. Again, the role-model effect cannot explain why top-performing female students perform worse when matched with female teachers. A more plausible explanation might be that female teachers target average/low-performing female students (Egalite, Kisida and Winters, 2015). It is possible that female teachers design their lectures with average female students in mind, and therefore those female students perform better from gender matching.

(c) Gains from matching in math, statistics, and econometrics are significantly lower than all other courses

Table IV shows that the gains from matching in math-intensive courses such as math, statistics, and econometrics are significantly lower than in the remaining courses. Earlier literature suggests that there is a lack of female role models in math-intensive courses at the college level (Freeman, 2004). Thus, if the role-model effect is the key mechanism, gains from matching should be higher in math courses, in which the dearth of female role models is greater than in social science courses. The literature on racial interaction suggests that black teachers help black students by incorporating *culturally relevant pedagogy* (Ladson-Billings, 1995) and teaching *hidden curricula* (Foster, 1990).[16] To the extent that employing culturally relevant pedagogy or teaching hidden curricula is easier in social science courses, the teacher-effectiveness channel can explain the lower gains from matching in math-intensive courses.

5.3.2 A Formal Test to Distinguish the Role-Model Effect from Teacher Effectiveness

To formally distinguish between the role-model effect and teacher-effectiveness mechanism, we use the testable implications of the model proposed by Gershenson et al. (2018), which distinguishes between these two mechanisms in the case of same-race student-teacher interaction. They posit an education production function with two inputs and two periods of investment. The two inputs—student investment (effort) and teacher quality—determine educational achievement. Guided by the pedagogical literature, they allow black teachers to raise the achievement levels of black students more than white teachers to capture the teacher-effectiveness channel (Ladson-Billings, 1995; Foster, 1990). In order to incorporate the role-model effect, Gershenson et al. (2018) assume black students have incorrect prior beliefs about the education production function by underestimating the returns to effort. They argue that black teachers are role models in the sense of providing "signals" to students about the true return to effort that leads students to update their beliefs about the education production function and increase their effort (investment). Again, they are guided by earlier studies that model the role-model effect as shifting beliefs about a parameter in a

[16] See Gershenson et al. (2018) for a detailed review.

production function (Steele, 1997; Cunha, Elo and Culhane, 2013; Papageorge, Gershenson and Kang, 2020).

According to the model of Gershenson et al. (2018), the effectiveness channel suggests that two black teachers would have a larger impact than one. In contrast, the role-model effect suggests that a second teacher may have a smaller effect if crucial information regarding returns to human capital has been transmitted by the first black teacher. Their model formalizes this insight and generates the following testable implications using *dosage models*, i.e., the marginal effect of having additional female teachers: If the teacher-effectiveness channel is the key mechanism, we should not observe diminishing returns to additional female teachers. On the other hand, if the role-model effect is the main mechanism, we would expect to observe diminishing returns.

I use the testable implications of Gershenson et al. (2018) about the marginal effect of being assigned an additional female teacher to distinguish between these two potential mechanisms. To this end, I run a variant of equation (4) to include the square of the percentage of female teachers and its interaction with the female student dummy. The results in Table X show that the interaction of the square of the proportion of female teachers with the female student dummy is positive and statistically significant for cumulative grade point average (CGPA), and the likelihood of enrolling in an economics master's program. I do not find any significant diminishing returns to having additional female teachers on other longer-term outcomes, such as the likelihood of graduating with the regular cohort or the number of years repeated. Thus, according to the testable implications described earlier, teacher effectiveness is an important channel through which female students benefit from having female teachers.

6 Conclusion

This study is the first to estimate the causal effects of female student-teacher interactions at the college level in the context of developing countries. My focus on mandatory courses, along with the limited ability of students to select teachers, allows me to attain a near-random setting to estimate the causal effects. I find that matching female students with female teachers significantly improves academic performance. Further, exposure to more female teachers improves female students' medium-term outcomes—such as cumulative GPA and the likelihood of finishing the degree with the regular cohort—and longer-term outcomes, such as the likelihood of enrolling in an economics master's program. I show that this improvement is not driven by female teachers' bias toward female students when grading.

My findings are consistent with previous studies that find same-gender teacher assignment matters at the college level. However, earlier studies that estimate female student-teacher interaction at the college level are in the context of developed countries, and to the best of my knowledge none focus on developing countries. The question of the role of female teachers in bridging the gender gap in educational attainment is much more salient in developing countries, in which stereotypes and prejudice are much more prevalent (Jayachandran, 2015) and gender gaps in college enrollment and attainment much larger (Ilie and Rose, 2016). In this paper, I document that—as is the case in developed countries—gender interaction also matters in developing countries, despite the differences in context. Prior research indicates that female students gain from gender matching in schools in developing countries (Muralidharan and Sheth, 2016; Rawal, Kingdon et al., 2010). I show that in a high-stakes setting such as college, this interaction continues to be beneficial for female students.

I document several other important results in my paper. First, gains from female student-teacher matching are higher in social science courses than in math-intensive courses. Second, I show that the interaction effects are higher for teachers with more experience. Third, below-median female students benefit more from having female teachers. Finally, exposure to female teachers solely in the first semester does not improve female students' long-term outcomes. Rather, continuous exposure to female teachers is what matters for improving female students' long-term academic achievement.

I also provide suggestive evidence that female teacher effectiveness in teaching female students is an important channel through which same-sex teacher assignment improves female students' academic achievement, in addition to the oft-cited role-model effect (Carrell, Page and West, 2010; Hoffmann and Oreopoulos, 2009; Bettinger and Long, 2005). This finding has important policy implications. If the role model is the key mechanism, as suggested by prior literature, then the only way to benefit female students is to appoint more female teachers in college. On the other hand, if teachers' experience or qualifications play a role, which suggests that pedagogical style is at play, then female students can benefit if college authorities acquaint male teachers with relevant pedagogical techniques that render female teachers more effective in teaching female students. Further, educating male teachers about stereotype threats can reduce their gender stereotypes (Hill, Corbett and St Rose, 2010) and cause them to be more accessible to female students.

Finally, I show that the benefit of gender matching in college is not limited to short- and medium-term outcomes but also extends to longer-term outcomes, such as the probability of enrolling in an economics graduate program. This finding is of significant relevance. Female underrepresentation in graduate economics courses has attracted great attention recently (Lundberg and Stearns, 2019). My results suggest that appointing more qualified female teachers can be an important tool for reducing the gender gap in a male-dominated field such as economics without affecting the performance of males.

References

Ambady, Nalini, Margaret Shih, Amy Kim, and Todd L Pittinsky. 2001. "Stereotype susceptibility in children: Effects of identity activation on quantitative performance."

Psychological science, 12(5): 385–390.

Aronson, Joshua, Diane M Quinn, and Steven J Spencer. 1998. "Stereotype threat and the academic underperformance of minorities and women." In *Prejudice*. 83–103. Elsevier.

Attiyeh, Gregory, and Richard Attiyeh. 1997. "Testing for bias in graduate school admissions." *The Journal of Human Resources*, 32(3): 524.

Bar, Talia, and Asaf Zussman. 2012. "Partisan grading." *American Economic Journal: Applied Economics*, 4(1): 30–48.

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(2): 152–157.

Blank, Rebecca M. 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review." *American Economic Review*, 81(5): 1041–1067.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre. 2020. "Do female role models reduce the gender gap in science? Evidence from French high schools."

Burgess, Simon, and Ellen Greaves. 2013. "Test scores, subjective assessment, and stereotyping of ethnic minorities." *Journal of Labor Economics*, 31(3): 535–576.

Carlana, Michela. 2019. "Implicit stereotypes: Evidence from teachers' gender bias." *The Quarterly Journal of Economics*, 134(3): 1163–1224.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap*." *The Quarterly Journal of Economics*, 125(3): 1101–1144.

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(1): 236–264.

Croizet, Jean-Claude, and Theresa Claire. 1998. "Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds." *Personality and Social Psychology Bulletin*, 24(6): 588–594.

Cunha, Flávio, Irma Elo, and Jennifer Culhane. 2013. "Eliciting maternal expectations about the technology of cognitive skill formation." National Bureau of Economic Research.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *The Review of Economics and Statistics*, 86(1): 195–210.

Egalite, Anna J, Brian Kisida, and Marcus A Winters. 2015. "Representation in the classroom: The effect of own-race teachers on student achievement." *Economics of Education Review*, 45: 44–52.

Falch, Torberg, and Linn Renée Naper. 2013. "Educational evaluation schemes and gender gaps in student achievement." *Economics of Education Review*, 36: 12 – 25.

Filer, Randall K. 1983. "Sexual differences in earnings: The role of individual personalities and tastes." *Journal of Human Resources*, 18(1).

Foster, Michele. 1990. "The politics of race: Through the eyes of African-American teachers." *Journal of education*, 172(3): 123–141.

Freeman, Catherine E. 2004. "Trends in Educational Equity of Girls & Women: 2004. NCES 2005-016." *National Center for Education Statistics*.

Gershenson, Seth, Cassandra M. D Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge. 2018. “The Long-Run Impacts of Same-Race Teachers.” National Bureau of Economic Research Working Paper 25254.

Goldin, Claudia, and Cecilia Rouse. 2000. “Orchestrating Impartiality: The Impact of ”Blind” Auditions on Female Musicians.” *American Economic Review*, 90(4): 715–741.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. “The homecoming of American college women: The reversal of the college gender gap.” *Journal of Economic perspectives*, 20(4): 133–156.

Hanna, Rema N, and Leigh L Linden. 2012. “Discrimination in grading.” *American Economic Journal: Economic Policy*, 4(4): 146–68.

Hill, Catherine, Christianne Corbett, and Andresse St Rose. 2010. *Why so few? Women in science, technology, engineering, and mathematics.* ERIC.

Hinnerich, Björn Tyrefors, Erik Höglin, and Magnus Johannesson. 2011. “Are boys discriminated in Swedish high schools?” *Economics of Education review*, 30(4): 682–690.

Hoffmann, Florian, and Philip Oreopoulos. 2009. “A Professor like Me: The Influence of Instructor Gender on College Achievement.” *The Journal of Human Resources*, 44(2): 479–494.

Ilie, Sonia, and Pauline Rose. 2016. “Is equal access to higher education in South Asia and sub-Saharan Africa achievable by 2030?” *Higher Education*, 72(4): 435–455.

Jayachandran, Seema. 2015. “The Roots of Gender Inequality in Developing Countries.” *Annual Review of Economics*, 7(1): 63–88.

Jones, Ethel B, and John D Jackson. 1990. “College grades and labor market rewards.” *The Journal of Human Resources*, 25(2): 253.

King, Sabrina Hope. 1993. "The limited presence of African-American teachers." *Review of Educational Research*, 63(2): 115–149.

Ladson-Billings, Gloria. 1995. "Toward a theory of culturally relevant pedagogy." *American educational research journal*, 32(3): 465–491.

Lavy, Victor. 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of Public Economics*, 92(10): 2083 – 2105.

Lindahl, Erica. 2016. "Are teacher assessments biased? Evidence from Sweden." *Education Economics*, 24(2): 224–238.

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in economics: Stalled progress." *Journal of Economic Perspectives*, 33(1): 3–22.

Luoto, Jill, Craig McIntosh, and Bruce Wydick. 2007. "Credit information systems in less developed countries: A test with microfinance in Guatemala." *Economic Development and Cultural Change*, 55(2): 313–334.

Muralidharan, Karthik, and Ketki Sheth. 2016. "Bridging education gender gaps in developing countries: The role of female teachers." *Journal of Human Resources*, 51(2): 269–297.

Neumark, David, and Rosella Gardecki. 1998. "Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics." *The Journal of Human Resources*, 33(1): 220–246.

Nguyen, Hannah-Hanh D, and Ann Marie Ryan. 2008. "Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence." *Journal of applied psychology*, 93(6): 1314.

Papageorge, Nicholas W, Seth Gershenson, and Kyung Min Kang. 2020. "Teacher expectations matter." *Review of Economics and Statistics*, 102(2): 234–251.

Rawal, Shenila, Geeta Kingdon, et al. 2010. "Akin to my teacher: Does caste, religious or gender distance between student and teacher matter? Some evidence from India." Department of Quantitative Social Science-UCL Institute of Education, University College London.

Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review*, 48(3): 515–530.

Spencer, Steven J, Claude M Steele, and Diane M Quinn. 1999. "Stereotype threat and women's math performance." *Journal of experimental social psychology*, 35(1): 4–28.

Steele, Claude M. 1997. "A threat in the air: How stereotypes shape intellectual identity and performance." *American psychologist*, 52(6): 613.

Vincent-Lancrin, Stéphan. 2008. "The reversal of gender inequalities in higher education: An on-going trend." *Higher education to*, 2030(1): 265–298.

Wise, David A. 1975. "Academic achievement and job performance." *The American Economic Review*, 65(3): 350–366.

Xie, Yu, and Kimberlee A Shauman. 2003. *Women in science.* Harvard university press.

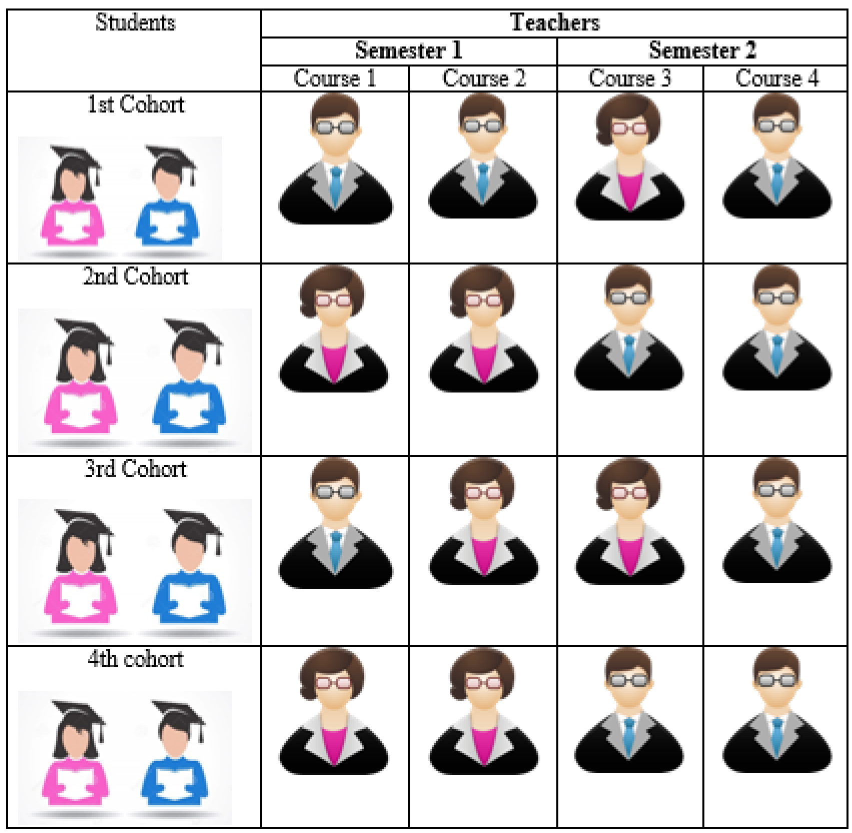

Figure 1: Hypothetical match between students and teachers across cohorts

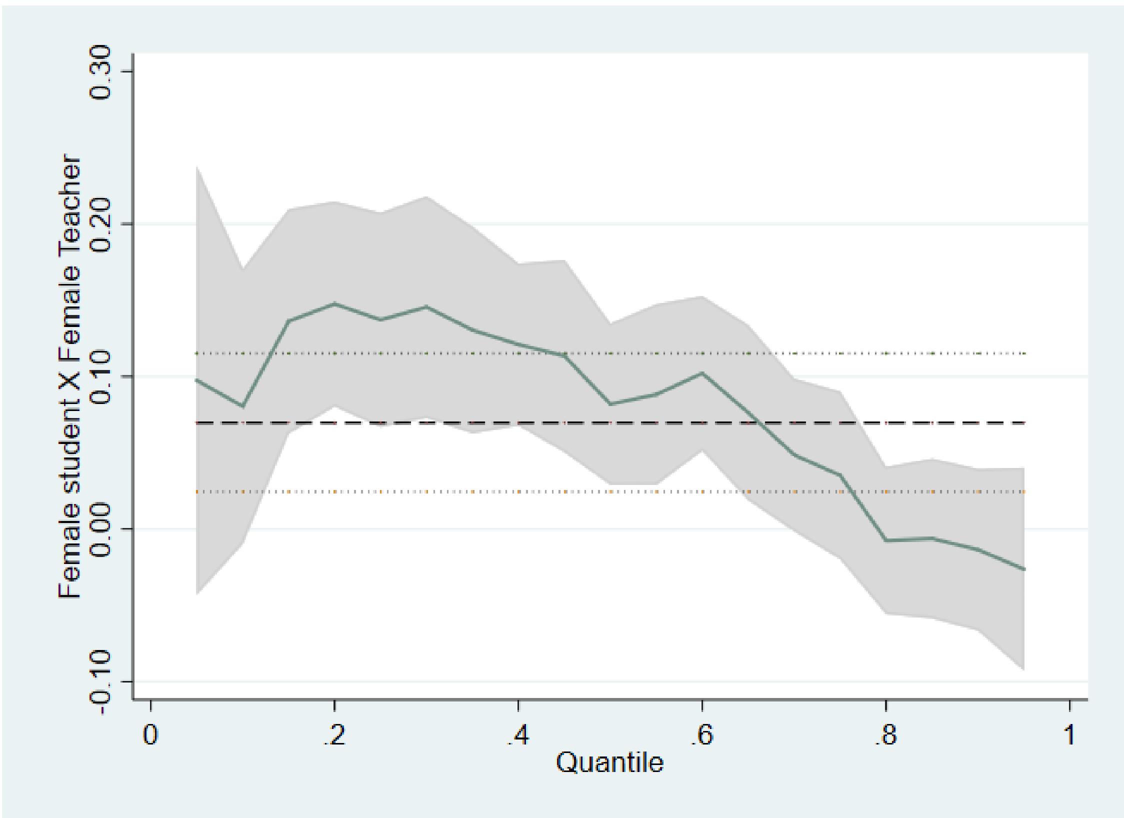

Figure 2: Estimates of the interaction effects at different quantiles of Normalized score
Notes: Estimates are from a quantile regression of student teacher interaction on normalized test scores.

Table I: Summary Statistics

	Number of observations	**Male**	**Female**
Student Profile			
Total students	1508	939 (62%)	569 (38%)
Average CGPA	1502	2.95	3.14
Average CGPA Conditional on Passing all courses	1280	3.18	3.27
Passed with Regular Batch	941	516 (55%)	425 (75%)
Number of years repeated			
Once	215	156	59
Twice	56	45	11
Thrice	6	4	2
Teacher Profile			
Total Teachers	66	46 (70%)	20 (30%)
Have a Terminal PhD Degree	30	20 (43%)	10 (50%)
Course offered			
No of courses taught by	402	246 (61%)	156 (39%)
Number of mandatory courses taught by	243	142 (58%)	101 (42%)
Number of courses taught by a instructor with a PhD Degree	212	133 (63%)	79 (37%)
Number of mandatory courses taught by a instructor with a PhD Degree	114	68 (60%)	46 (40%)
Number of first-year mandatory courses taught by	86	73 (85%	13 (15%)

Table II: Comparison of Mean Characteristics

	Male Teacher	**Female Teacher**	**Difference**	**P-value**	**Observation**
Panel A: Female Students					
Merit serial in entrance exam	215.6	210.9	4.73	0.42	13169
Merit score in entrance exam	147.4	147.5	-0.1	0.89	13169
High school degree was science	0.459	0.453	0.005	0.7	13169
Had quota	0.081	0.08	0.001	0.76	13169
HSC GPA	4.393	4.385	0.008	0.49	2987
SSC GPA	4.5	4.47	0.03	0.51	2987
Took math in HSC	0.333	0.317	0.016	0.56	3262
Took Statistics in HSC	0.141	0.147	-0.006	0.56	3262
Panel B: Male Students					
Merit serial in entrance exam	230.47	218.83	11.63	0.09	22438
Merit score in entrance exam	146.21	146.63	-0.43	0.56	22438
High school degree was science	0.49	0.48	0.01	0.23	22438
Had quota	0.068	0.065	0.003	0.36	22438
HSC GPA	4.211	4.21	0.001	0.93	5556
SSC GPA	4.44	4.43	0.01	0.54	5556
Took math in HSC	0.284	0.272	0.011	0.56	5844
Took Statistics in HSC	0.203	0.2	0.003	0.56	5844

Notes: Unit of observation is individual student by course. We have partial data (only two out of 11 cohorts) on the last four variables of each panel. HSC stands for Higher Secondary Exam, and SSC stands for Secondary School Exam. HSC and SSC exams are national-level exams that take place after 12 years and 10 years of schooling respectively. The P-values are adjusted for intra-classroom correlation.

Table III: Effect of Instructor's Gender on Students' Academic Performance

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Standardized test score						
Female teacher x Female student	0.0681**	0.0494*	0.0877***	0.0869***	0.0862***	0.0864***	0.0935***
	(0.0272)	(0.0260)	(0.0246)	(0.0248)	(0.0248)	(0.0248)	(0.0256)
Female student	0.199***	0.174***					
	(0.0203)	(0.0199)					
Female teacher	0.0282**	0.0152	0.0734***				
	(0.0109)	(0.0171)	(0.0163)				
Student characteristics at entry	No	Yes	No	No	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Teacher fixed effects	No	No	No	Yes	Yes	Yes	Yes
Course fixed effects	No	No	No	No	Yes	Yes	Yes
Teacher gender x Course fixed effects	No	No	No	No	No	Yes	Yes
Classroom fixed effects	No	No	No	No	No	No	Yes
Observations	31,613	27,853	31,613	31,613	31,613	31,613	31,613
R-squared	0.013	0.029	0.571	0.575	0.576	0.576	0.584

Notes: Unit of observation is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table IV: Effect of Instructor's Gender on Students' Academic Performance by Subject Type

	Dependent Variable: Standardized Score	
	Math, Statistics, and Econometrics	Other courses
VARIABLES	(1)	(2)
Female Teacher x Female Student	0.0596*	0.123***
	(0.0346)	(0.0364)
Student fixed effects	Yes	Yes
Teacher fixed effects	Yes	Yes
Course fixed effects	Yes	Yes
Teacher gender x Course fixed effects	Yes	Yes
Classroom fixed effects	Yes	Yes
Observations	11,333	20,262
R-squared	0.675	0.571

Notes: Unit of observation is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table V: Effect of Instructor's Gender on Long-term Outcomes

	Bachelor Degree			Master's degree	
	Cumulative grade point average	Passed with regular cohort	Number of years repeated	Enrolled	Cumulative grade point average
VARIABLES	(1)	(2)	(3)	(4)	(5)
Female student	-0.0787	-0.0214	0.124	-0.411*	0.210
	(0.131)	(0.0987)	(0.116)	(0.233)	(0.414)
% of female teachers in all mandatory courses	0.0253***	0.0141***	-0.00750***	0.0134***	0.00279
	(0.00186)	(0.00140)	(0.00165)	(0.00249)	(0.00971)
% of female teachers in all mandatory courses x Female student	0.00549*	0.00459**	-0.00585**	0.0114*	-0.00433
	(0.00291)	(0.00219)	(0.00258)	(0.00598)	(0.0106)
Observations	1,496	1,496	1,496	654	524
R-squared	0.239	0.193	0.085	0.354	0.513
Male student mean	2.951	0.550	0.279	0.748	3.129

Notes: Unit of Analysis is the individual student. The regression follows equation (4) as specified in the main text. Information on the Master's degrees was available only for the first four cohorts of students. All the regressions include cohort fixed effects. Standard errors are corrected for cohort-level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table VI: Effect of Instructor's Gender on Students' Academic Performance in Blind and Non-blind Exams

	Dependent Variable: Standardized Score	
	Non-blind exam	Blind exam
VARIABLES	(1)	(2)
Female Teacher x Female Student	0.0872***	0.0664**
	(0.0259)	(0.0286)
Student fixed effects	Yes	Yes
Teacher fixed effects	Yes	Yes
Course fixed effects	Yes	Yes
Teacher gender x Course fixed effects	Yes	Yes
Classroom fixed effects	Yes	Yes
Observations	31,930	31,571
R-squared	0.484	0.474

Notes: Unit of analysis is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table VII: Effect of Instructor's Gender on Students' Academic Performance with Blind Exam Interaction

VARIABLES	Dependent variable: Standardized test score (1)	(2)
Female Teacher x Female Student	0.0926***	0.0861***
	(0.0263)	(0.0259)
Blind Exam	-0.00886	
	(0.00875)	
Female Student x Blind Exam	0.000788	
	(0.0220)	
Female Teacher x Blind Exam	0.0116	0.00298
	(0.0125)	(0.0125)
Female Teacher x Female Student x Blind Exam	-0.0322	-0.0218
	(0.0310)	(0.0300)
Student fixed effects	Yes	Yes
Teacher fixed effects	Yes	Yes
Course fixed effects	Yes	Yes
Teacher gender x Course fixed effects	Yes	Yes
Classroom fixed effects	Yes	Yes
Student x Exam type Fixed effects	No	Yes
Observations	63,502	63,499
R-squared	0.461	0.484

Notes: Unit of analysis is student-course-exam type. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table VIII: Teachers' Qualifications and Gender Interaction

	Normalized test score		
Teachers' characteristics	Rank	Experience	Has a Ph.D. degree
	(1)	(2)	(3)
Panel A: All compulsory courses			
Female student x Female teacher	-0.0962*	-0.0458	0.0861**
	(0.0575)	(0.0420)	(0.0343)
Female student x Female teacher x Teacher Characteristics	0.0766***	0.0103***	0.0264
	(0.0215)	(0.00259)	(0.0489)
Observations	31,499	31,499	31,499
Panel B: Maths, Statistics, and Econometrics			
Female student x Female teacher	-0.188**	-0.125**	-0.0631
	(0.0888)	(0.0615)	(0.0710)
Female student x Female teacher x Teacher Characteristics	0.111***	0.0195***	0.217**
	(0.0394)	(0.00625)	(0.0911)
Observations	11,333	11,333	11,333
Panel C: Other courses			
Female student x Female teacher	-0.0339	0.0156	0.133***
	(0.0836)	(0.0591)	(0.0482)
Female student x Female teacher x Teacher Characteristics	0.0616**	0.00800**	-0.0301
	(0.0301)	(0.00335)	(0.0659)
Observations	20,148	20,148	20,148

Notes: Unit of analysis is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (teacher-course-cohort) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table IX: Students' Qualifications and Gender Interaction

Student Characteristics at entry	Normalized test score Merit Serial	Merit score	High school discipline is science	Admitted under quota
	(1)	(2)	(3)	(4)
Panel A: All compulsory courses				
Female student x Female teacher	0.0808***	0.118**	0.0739***	0.0836***
	(0.0248)	(0.0493)	(0.0250)	(0.0245)
Female student x Female teacher x Student Characteristics	3.13e-05	-0.000229	0.0313	0.0510
	(3.73e-05)	(0.000283)	(0.0298)	(0.0466)
Observations	31,613	31,613	31,613	31,613
Panel B: Maths, Science, and Statistics				
Female student x Female teacher	0.0523	0.117	0.0479	0.0456
	(0.0361)	(0.0756)	(0.0316)	(0.0338)
Female student x Female teacher x Student Characteristics	7.51e-06	-0.000460	0.0142	0.128*
	(6.24e-05)	(0.000486)	(0.0489)	(0.0687)
Observations	11,333	11,333	11,333	11,333
Panel C: Other courses				
Female student x Female teacher	0.109***	0.105	0.107***	0.116***
	(0.0350)	(0.0658)	(0.0353)	(0.0349)
Female student x Female teacher x Student Characteristics	4.22e-05	9.00e-05	0.0251	0.0154
	(4.54e-05)	(0.000358)	(0.0327)	(0.0626)
Observations	20,262	20,262	20,262	20,262

Notes: Unit of analysis is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are normalized test scores. Standard errors are corrected for classroom (teacher-course-cohort) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table X: Effect of Marginal Increase in Female Teachers on the Medium- and Longer-term Outcomes

VARIABLES	(1) CGPA	(2) Passes with regular batch	(3) Number of years repeated	(4) Enroll in masters program	(5) Masters CGPA
Female student	0.251	-0.279	0.0490	0.0777	2.702
	(0.253)	(0.188)	(0.225)	(0.382)	(8.987)
% of female teachers in all mandatory courses	0.0610***	0.0275***	0.00269	0.0442***	0.0913
	(0.00573)	(0.00425)	(0.00510)	(0.00945)	(0.315)
% of female teachers in all mandatory courses **squared**	-0.000508***	-0.000200***	-0.000150**	-0.000419**	-0.00113
	(7.61e-05)	(5.63e-05)	(6.77e-05)	(0.000179)	(0.00406)
% of female teachers in all mandatory courses x Female student	-0.0134	0.0154*	-0.00311	-0.0257	-0.135
	(0.0114)	(0.00846)	(0.0102)	(0.0226)	(0.461)
% of female teachers in all mandatory courses **squared** x Female student	0.000241*	-0.000110	-3.22e-05	0.000652*	0.00174
	(0.000132)	(9.75e-05)	(0.000117)	(0.000368)	(0.00588)
Observations	1,463	1,463	1,463	629	507
R-squared	0.259	0.214	0.095	0.171	0.030
Male student mean	2.945	0.556	0.290	0.755	3.124

Notes: Unit of Analysis is the individual student. The regression is a modified version of equation (4) as specified in the main text. Information on the Master's degrees was available only for the first four cohorts of students. All the regressions include cohort-fixed effects. Standard errors are corrected for Cohort-level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table XI: Teachers' Qualifications and Gender Interaction

	Dependent variable: Standardized test score							
	Rank			Teaching experience			PhD status	
VARIABLES	Lecturer	Assistant/ Associate Professor	Professor	<10 years	10-20 years	>20 years	Doesn't have a PhD degree	Has a PhD degree
Female Teacher x Female Student	-0.126**	0.144***	0.208***	-0.0403	0.195***	0.245***	0.0926***	0.120***
	(0.0596)	(0.0357)	(0.0562)	(0.0465)	(0.0400)	(0.0754)	(0.0353)	(0.0384)
Student fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Course fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher gender x Course fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,448	12,836	10,107	15,322	7,164	8,794	17,207	14,246
R-squared	0.667	0.635	0.583	0.633	0.628	0.630	0.623	0.578

Notes: Unit of analysis is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table XII: Effect of Instructor's Gender on long term-outcomes

	Bachelor Degree			Master's degree	
	Cumulative grade point average	Passed with regular cohort	Number of years repeated	Enrolled	Cumulative grade point average
VARIABLES	(1)	(2)	(3)	(4)	(5)
Female student	0.208***	0.237***	-0.169***	0.0268	0.0992**
	(0.0493)	(0.0362)	(0.0416)	(0.0456)	(0.0493)
% of female teachers in 1st year mandatory courses	0.00282	0.00486***	-4.94e-05	0.00176	0.00475
	(0.00223)	(0.00164)	(0.00188)	(0.00290)	(0.00405)
% of female teachers in 1st year mandatory courses x Female student	-0.00236	-0.00345**	0.00166	0.000197	-0.00415
	(0.00232)	(0.00171)	(0.00196)	(0.00239)	(0.00261)
Observations	1,485	1,485	1,485	650	524
R-squared	0.120	0.120	0.062	0.317	0.516
Male student mean	2.951	0.550	0.279	0.748	3.129

Notes: The unit of analysis is the individual student. Each column presents estimated coefficients from a linear fixed effect regression for different long-term outcome variables. Information on the Master's degree was available only for the first four cohorts of students. All the regressions include cohort fixed effects. Standard errors are corrected for batch-level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Appendix Figures and Tables

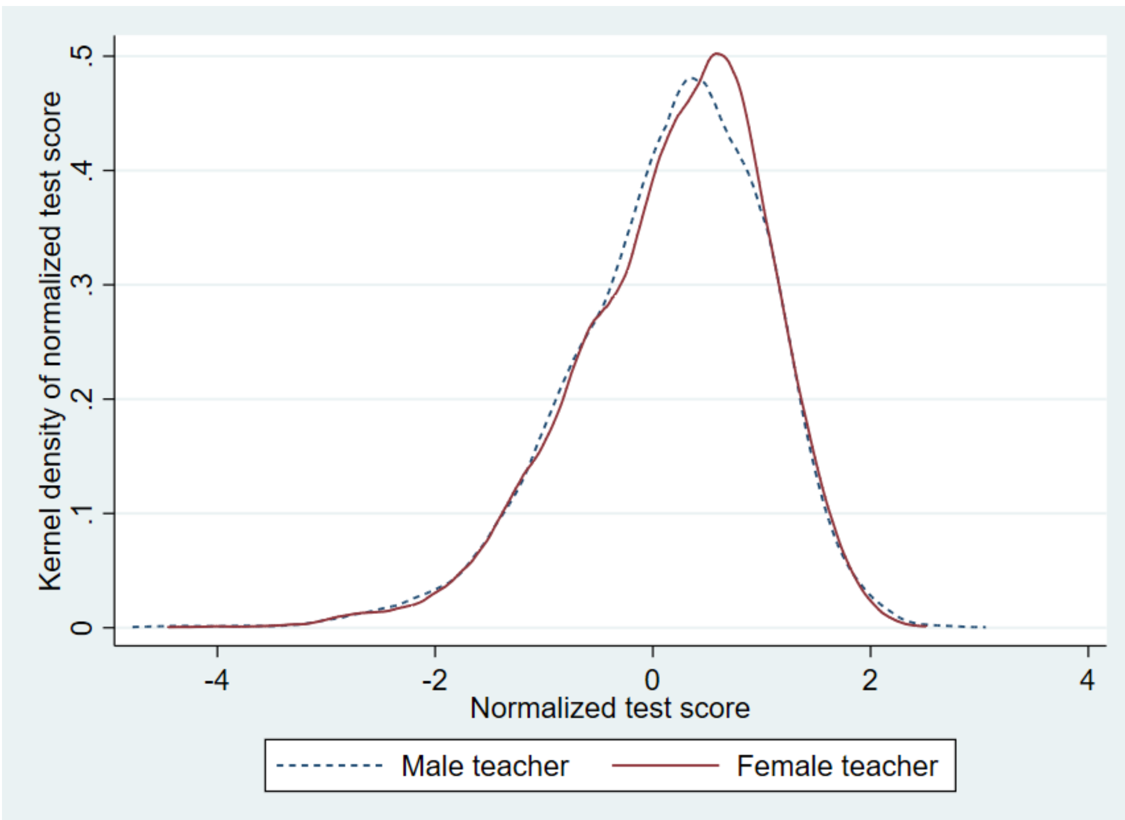

Figure A.1: Distribution of female students' test score by the gender of the teacher

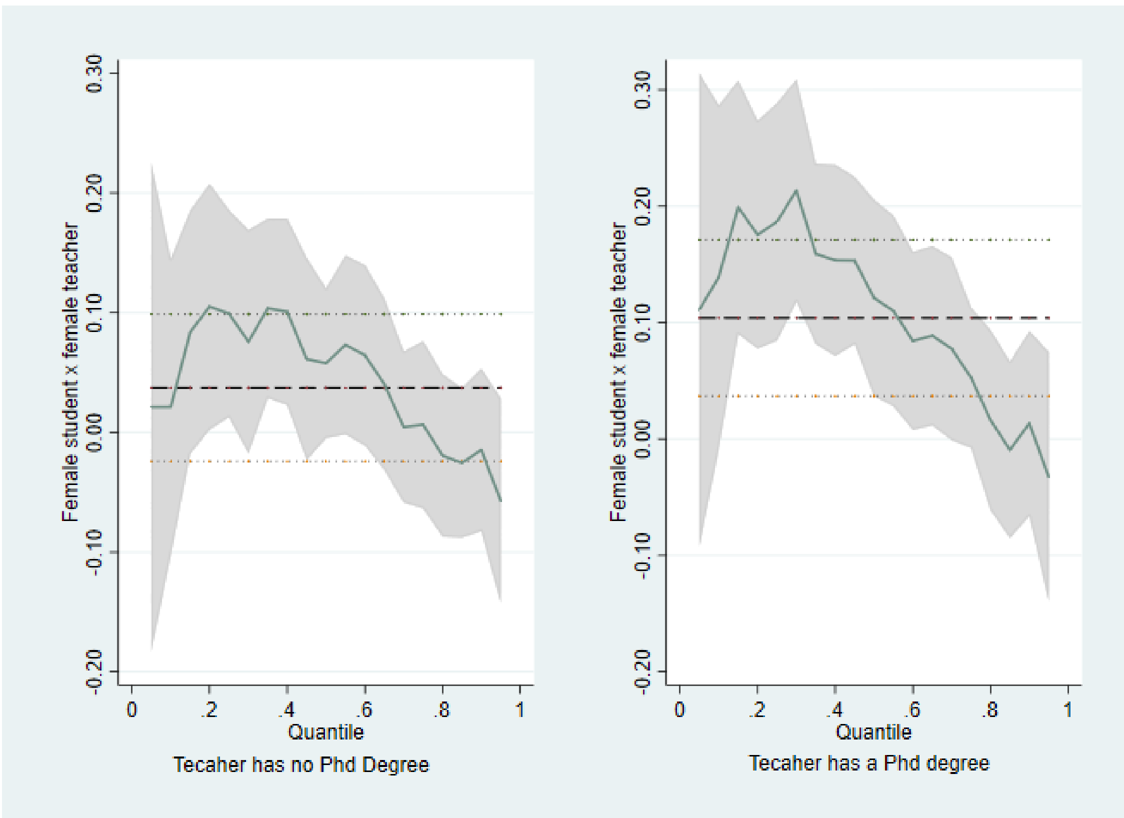

Figure A.2: Quantile Treatment Effect: By PhD Status of the Instructor

Table A.1: Grading System Used in the Department of Economics

Numerical Marks	**Grade**	**Grade Point**
80 to 100	A+	4.00
75 to 79	A	3.75
70 to 74	A-	3.50
65 to 69	B+	3.25
60 to 64	B	3.00
55 to 59	B-	2.75
50 to 54	C+	2.50
45 to 49	C	2.25
40 to 44	D	2.00
<40	F	0.00

Note: Minimum GPA required for promotion is 2.00

Table A.2: Effect of Female Student Dummy on the Proportion of Female Teachers in Compulsory Courses in Different Semesters

	Proportion of female teachers in mandatory courses							
VARIABLES	1st semester (1)	2nd semester (2)	3rd semester (3)	4th semester (4)	5th semester (5)	6th semester (6)	7th semester (7)	8th semester (8)
Female student	-0.00446	-0.00205	-0.00477	0.00485	0.00205	0.00349	-0.000844	0.00106
	(0.00477)	(0.00769)	(0.00553)	(0.00281)	(0.00472)	(0.00525)	(0.00403)	(0.00499)
Constant	0.150***	0.168***	0.759***	0.602***	0.577***	0.731***	0.316***	0.373***
	(0.00182)	(0.00297)	(0.00218)	(0.00111)	(0.00191)	(0.00214)	(0.00169)	(0.00209)
Observations	1,483	1,429	1,265	1,239	1,097	1,082	928	691
R-squared	0.789	0.805	0.912	0.892	0.921	0.964	0.967	0.970

Notes: The unit of analysis is the individual student. Each column presents estimated coefficients from a linear fixed effect regression of a dummy variable for female students in semester t on the proportion of teachers in compulsory courses in semester t. Each regression includes cohort fixed effect. Standard errors are corrected for cohort-level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table A.3: Effect of Female Student's Academic Performance on the Proportion of Female Teachers in Compulsory Courses in the Subsequent Semester

	Proportion of female teachers in mandatory courses						
VARIABLES	2nd semester (2)	3rd semester (3)	4th semester (4)	5th semester (5)	6th semester (6)	7th semester (7)	8th semester (8)
CGPA of female students in the preceding semster	0.000762	0.00670	-0.00647	-0.00480	0.00854	0.00861	0.0112
	(0.0154)	(0.0154)	(0.00415)	(0.00667)	(0.0123)	(0.00775)	(0.00802)
Constant	0.163***	0.734***	0.627***	0.597***	0.707***	0.282***	0.335***
	(0.0485)	(0.0500)	(0.0134)	(0.0219)	(0.0406)	(0.0257)	(0.0264)
Observations	551	499	492	444	441	389	289
R-squared	0.877	0.929	0.926	0.972	0.988	0.991	0.992

Notes: Unit of analysis is individual student. Each column presents estimated coefficients from a linear fixed effect regression of female students' academic performance in semester $t-1$ on the proportion of teachers in compulsory courses in semester t . Each regression includes cohort fixed effect. Standard errors are corrected for cohort level clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table A.4: Effect of Instructor's Gender on Students' Academic Performance in Single Blind and Double blind Exams

	Dependent Variable: Standardized Score	
	Double-blind exam	Single-blind exam
VARIABLES	(1)	(2)
Female Teacher x Female Student	0.0745***	0.0617**
	(0.0283)	(0.0274)
Student fixed effects	Yes	Yes
Teacher fixed effects	Yes	Yes
Course fixed effects	Yes	Yes
Teacher gender x Course fixed effects	Yes	Yes
Classroom fixed effects	Yes	Yes
Observations	14,050	17,520
R-squared	0.451	0.502

Notes:Unit of analysis is individual student-course. Each column presents estimated coefficients from a linear fixed effect regression for students' academic performance. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.

Table A.5: Effect of Instructor's Gender on Students' Academic Performance by Entry Score Quantile

	Dependent Variable: Standardized Score			
VARIABLES	1st Quantile (1)	2nd quantile (2)	3rd quantile (3)	4th quantile (4)
Female Teacher x Female Student	0.0731*	0.117***	0.0561	0.0614*
	(0.0427)	(0.0364)	(0.0458)	(0.0335)
Student Fixed Effect	Yes	Yes	Yes	Yes
Cohort-Semester Fixed Effect	Yes	Yes	Yes	Yes
Teacher Fixed Effect	Yes	Yes	Yes	Yes
Course Fixed Effect	No	Yes	No	Yes
Observations	6,673	6,761	7,018	7,390
R-squared	0.526	0.564	0.600	0.644

Notes: Unit of analysis is student-course. Each column presents estimated coefficients from a linear fixed effect regression of students' academic performance at different entry score quantiles. Dependent variables are standardized scores. Standard errors are corrected for classroom (cohort-semester-course) clustering and are presented in parentheses. *** indicates significance at 1, ** at 5, and * at 10 percent level.