The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea

Jaegeum Lim*
Texas A&M University

Jonathan Meer
Texas A&M University
and NBER

May 17, 2016

Abstract

Gender disparities in academic performance may be driven in part by the interaction of teacher and student gender, but systematic sorting of students into classrooms makes it difficult to identify causal effects. We use the random assignment of students to Korean middle school classrooms and show that the female students perform substantially better on standardized tests when assigned to female teachers; there is little effect on male students. We find evidence that teacher behavior drives the increase in female students' achievement.

1 Introduction

Gender gaps in academic performance, with girls generally outperforming boys in language arts and boys generally outperforming girls in math, have persisted despite decades of effort to close them. Understanding the causes of these gaps is crucial, especially at younger ages, as they may lead to gender differences in later course-taking, occupational choices, and labor market outcomes (Lavy and Sand, 2015).

One possible source of gender-based disparities is whether a student and a teacher share the same gender. These gender interactions may affect academic performance through changes in the behavior of both parties, that is, through student- or teacher-centered mechanisms. Role-model effects, an example of the former, predict that students will be more engaged in study when they are taught by the same-gender teacher (Dee, 2007). As an example of the latter, a teachers might assign less difficult homework questions to girls if he or she believes that girls are less capable in math than boys (Jones and Dindia, 2004).

The primary threat to identifying the causal effect of teacher-student gender matches is the nonrandom sorting of students that typifies classroom assignment

---

in most contexts. For instance, students with a lower propensity to achieve academically may be more likely to be assigned to a teacher of a particular gender. Beginning with Dee (2007), the standard approach in this literature has been to use student interactions with multiple teachers across different subjects. By using estimates including student fixed effects, unobserved student characteristics that are correlated with student quality and teacher gender will not bias estimation. Dee uses the fact that the National Education Longitudinal Survey of 1988 surveys two teachers for every student to estimate within-student teacher-gender effects; he finds evidence of substantial positive impacts on academic achievement of being assigned to a teacher of the same gender. Moreover, Dee uses subjective evaluations of both teacher and student perceptions to show that students are less likely to be seen as disruptive when evaluated by a teacher of the same gender, and more likely to report interest in that academic subject. Using a different approach, Muralidharan and Sheth (2016) exploit panel data from India – in particular, schools with only one classroom per grade, in which there can be no sorting of students. They find that female primary school students perform significantly better with female teachers, with no impact of teacher gender on male students.

On the other hand, several studies find no effect of teacher gender. Holmlund and Sund (2008) use Swedish secondary-school panel data and identify the impact of same-gender teachers using teacher turnover. Once they control for subject-specific gender effects, they find no impact of gender matching on student performance. Cho (2012) uses math and science test score data from 15 OECD countries and shows that there is no significant effect of teacher-student gender matching in eight of these countries, including the United States. Most recently, Paredes (2014) examines role model and teacher bias effects with data from Chile, finding small but statistically significant gender-matching effects for girls and no effects for boys, as well as suggestive evidence that role model effects drive the result.

However, this within-student estimation approach – even when including *teacher* fixed effects – is insufficient if students and teachers are systematically matched on characteristics correlated with gender. For instance, suppose female students who would benefit relatively more from having a female teacher are more likely to be assigned to female teachers who, themselves, are better role models for female students. In this case, a positive student-teacher gender interaction effect reflects sorting. As Dee notes, "the internal validity of such within-student comparisons could still be compromised by the nonrandom sorting by students with subject-specific propensities for achievement and by unobserved teacher and classroom traits correlated with gender." He finds some such evidence in the NELS:88 with a number of indirect tests, particularly in the assignment of female math teachers. Other studies lacking random assignment must indirectly show that the identification strategy holds; for instance, Paredes (2014) uses previous-year's test scores to control for achievement propensity.

For identification, we exploit a unique feature of secondary education in South Korea: the random assignment of students into a classroom, where students remain throughout the school day. We provide evidence for our identifying

2

assumption in a number of ways: first, as-good-as-random assignment of students to classrooms is a strict policy in South Korea. We confirm that schools follow this policy by surveying a large number of them on the topic. We also show that assignment to classrooms within a school is uncorrelated with observable characteristics. Furthermore, students who are assigned to same- and opposite-gender teachers look similar in their observable characteristics. Finally, our results do not differ when additional controls, student fixed effects, or teacher fixed effects are included, as one would expect if assignment is truly random.

Our reliance on random assignment obviates potential sorting issues that have been a major concern in previous work. In this way, our approach is most similar to two previous papers. Antecol et al. (2015) exploit the random assignment of students in an experiment testing the efficacy of Teach for America, a program that trains and places high-achieving new teachers at disadvantaged schools. They find that female elementary school students with female teachers perform *worse* than those with male teachers. However, this negative effect disappears for female teachers with stronger math backgrounds. At the higher education level, Carrell, Page and West (2010) use random assignment of cadets at the United States Air Force Academy to compulsory math and science courses and show that female professors significantly reduce the gender gap in performance for female students.[1]

Our study is also unique in that we provide more recent evidence from an age group similar to that studied in Dee (2007) and, importantly, our empirical setting is a culture with somewhat different gender norms than many of those previously studied. South Korea is ranked 39th of 57 countries in its residents' attitudes towards gender equality, much lower than the countries studied in the analyses above: 29th for Chile, 18th for the United States, and 2nd for Sweden (Brandt, 2011).[2]

Results show that female students' performance is positively influenced by having a female teacher, but that there is little same-gender teacher effect for males. The overall increase in the female-male performance gap of about a tenth of a standard deviation is comparable in size to those found by Dee (2007) and Carrell et al. (2010). Unlike our findings, though, both of those papers find that the impact is divided about evenly between reduced performance by males and increased performance by females. Our effect is similar in magnitude to an increase of one standard deviation in teacher quality (Chetty et al., 2014).

The impacts are primarily concentrated in mathematics and English language scores, as compared to Korean language scores. We also provide some suggestive evidence that teacher-centered mechanisms are behind these impacts,

---

[1] Other evidence on gender-matching effects on student grades, course-taking, and persistence in colleges is mixed; see, for example, Canes and Rosen (1995), Bettinger and Long (2005), and Hoffmann and Oreopoulos (2009).

[2] Our work is also related to the literature on the impact of single-sex schools. Park et al. (2013) find significant positive impacts of single-sex schooling using random assignment in South Korea, while Jackson (2012) exploits the nature of rules-based school assignment in Trinidad and Tobago and finds little effect for most students.

with female students reporting that their female teachers are more likely to encourage them and to give them an equal opportunity to express themselves.

2 Data

We use cross-sectional data collected by the Korean Educational Development Institute (KEDI) in July 2004, at the end of the first semester of middle school in South Korea. The target schools, covering 6.8% of the relevant population in South Korea in 2004, were selected by proportionate stratified random sampling. Our initial sample consists of 197 schools, 777 Korean, English, and mathematics teachers linked to surveyed classrooms, 14,372 students, and 11,944 parents. Thirty-five of the schools had all-female students and 35 were all-male; 84 classrooms are single-sex within 127 coed schools.[3] Restricted-use data provided by KEDI allows us to link students to classrooms.

In addition to an extensive set of questions, students' responses were linked to their scores on the Student Achievement Test administered by the Seoul Metropolitan Office of Education (SMOE). Students in the sample were tested at the beginning of the second semester of ninth grade in three courses: Korean language, English language, and mathematics; 12,363 students' test results were collected.[4]

The teacher questionnaire includes information on teachers' classroom assignments, which we use to link students with their subject teachers. Beginning with 37,034 student-subject combinations with test score information, we first drop 6,033 observations without classroom or teacher information. Of these, 42 observations from 14 students have missing classroom information and 5,991 observations from 224 classrooms do not have teacher information due to non-response by teachers, reducing the number of teachers in the sample to 777 and the number of students to 12,305.[5] For our primary sample, we also drop 6,442 observations for students with multiple subject teachers, for which we could not make a student-teacher match representing just one student and one teacher. We also show results including these observations, which are unchanged from those excluding them. This results in 24,489 student-teacher pairings representing 11,659 students and 502 teachers. Among them, 33% of observations correspond to a female student with a female teacher, 16% are a female student with a male teacher, 32% are a male student with a female teacher, and the remaining 19% are male students with male teachers.

---

[3]As discussed below, excluding single-sex schools or single-sex classrooms does not change our results.

[4]This exam is administered to 9th graders in Seoul every September; these students would have taken the test regardless. Students living outside of Seoul but in the KEDI sample took the same exam on the same day.

[5]A concern is that teacher non-response could somehow be correlated with their impact on students of different genders. While we cannot completely exclude this possibility, students dropped from the sample due to teacher non-response have similar test scores ($p = 0.77$) as those remaining in the sample. There were also no statistically significant differences in the other student characteristics we examined.

4

2.1 Classroom assignment in South Korea

2.1.1 Student assignment

Elementary school graduates in South Korea are randomly assigned to middle schools within their district.[6] At the beginning of each academic year (March 1st), middle school students in South Korea are assigned a classroom where they remain throughout a school day, and where each subject teacher visits to present a lesson. Be it private or public, schools in South Korea use some form of random assignment to classrooms due to both strong social norms and government policies (Kang, 2007). The most common approach is to order students by their academic performance in the previous year and assign them across classrooms. As an example, the top ranked student would be assigned to the first classroom, the second-ranked student assigned to the second classroom, and so on.[7] To confirm this point, we surveyed local Offices of Education on schools' rules for classroom assignment for the 197 schools in our sample.[8] All but one of the 180 responding schools with more than one classroom per grade reported that they used this method of classroom assignment, with the sole exception being a school that used alphabetical order of names to assign students.

2.1.2 Teacher assignment

Even with random assignment of students to classroom units, the internal validity of our approach is threatened if teachers are systematically assigned to those classrooms in a way that is related to their gender. For example, female teachers might be assigned to classrooms that, by chance, have students with less-involved parents. There are no written government guidelines on teacher assignment; we interviewed a number of current teachers and principals to gain insight into the process. First, homeroom teachers are assigned, either by lottery or a committee, to a particular classroom. These teachers, who teach a subject themselves, are responsible for discipline, taking attendance at the start of the day, and overseeing study halls before and after school. Subject teachers' classroom assignments are generally determined in an ad hoc way that is unrelated to student or teacher characteristics. For example, one subject teacher may take odd-numbered classrooms while the other takes even-numbered ones.

---

[6]Since 1996, students in districts whose superintendents allow it are permitted to list several preferred schools. They are entered into a lottery for each school on their preferred list (Korea Legislation Research Institute, 2011).

[7]Kang (2007) uses this same random assignment feature and a different data set on the performance of Korean students to examine peer effects. As mentioned above, Park et al. (2013) examine the effect of single-sex education on college-going behavior using data from Seoul, in which students are not allowed to list preferred schools. Lee et al. (2014) examine schools in the Seoul metropolitan area to study the effects of single-sex versus co-educational schooling on academic performance.

[8]Note, of course, that schools were responding eleven years after the KEDI survey was conducted. In recent years, the Korean education system has shifted somewhat from its original strictly egalitarian approach, so it seems quite likely that these as-good-as-random practices were in place in 2004. See, for example, Byun and Kim (2010), who discuss increased use of ability tracking in South Korea over the past decade.

We surveyed the schools in our sample on these policies as well, with 141 of 153 responding schools reporting that they assign subject teachers without considering student or teacher characteristics. The remaining 12 schools reported considering teachers' characteristics, such as experience, in making the assignment; our results are unchanged by excluding these schools, and we once again note that we conducted our survey eleven years after our data were collected. In Section 3.1, we further examine whether random assignment holds in our data based on students' and teachers' observable characteristics.

3 Methodology

3.1 Tests of Random Assignment

3.1.1 Resampling Techniques

While the institutional setting we study is clear that students are randomized across classrooms without respect to teacher gender, we also provide empirical evidence to support our identification strategy. We begin by following Carrell and West (2010), Lehmann and Romano (2005), and Good (2006) in using resampling techniques to test the randomness of teacher and student matching in terms of student's observable characteristics. First, for each classroom within a school, we randomly draw 10,000 synthetic classrooms of the same size from the sample of all students in the school, without replacement. We do so for each of the three subjects – Korean, English, and mathematics – and for each of six variables (indicator variables for student being male, parents being married, father with BA degree or higher, mother with BA degree or higher, having housing ownership, and having Internet access at home). Then, for each subject and characteristic combination, and for each classroom within a school, we calculate the number of students with the characteristic within a classroom.[9] We obtain an empirical p-value, namely, the proportion of the 10,000 resampled classrooms with fewer students with the characteristic (e.g. male) within the observed classroom.

Under random assignment, any p-value will be observed with equal probability; we therefore expect the empirical p-values to be uniformly distributed. We test whether the distribution of the empirical p-values for each subject and characteristic combination is uniform using Kolmogorov-Smirnov and χ^2 goodness-of-fit tests. Table 1A presents the results of this exercise, aggregating results over school subject for brevity. Overall, we reject random assignment for 34 out of 1,942 (1.8%) school-by-subject-by-characteristic test statistics[10] at 5 percent level in the Kolmogorov-Smirnov test and 71 of 1,942 (3.7%) test statistics using

---

[9]Carrell and West (2010) use the sums of SAT scores or academic composite to obtain empirical p-values. Similarly, we sum the indicator variables.

[10]For some school by subject by characteristic combinations, test statistics cannot be calculated because of missing variables, only a single classroom from the school remaining in the sample, or the school itself being single-sex for the student gender characteristic.

the χ^2 goodness of fit test. Therefore, we do not find evidence of nonrandom assignment of students into classrooms by observable characteristics.

We also check the random assignment of teachers with respect to student's observable characteristics. For each characteristic, we regress the empirical p-values on a set of teacher characteristics, controlling for subject and school fixed effects. The results, in Table 1B, show that only one of thirty coefficients is statistically significant at the 5 percent level. Therefore, there is little evidence of nonrandom assignment of teachers into classrooms with respect to student's observable characteristics.

Table 1: Randomness Check

	(1) Male	(2) Married Parents	(3) Dad BA Degree	(4) Mom BA Degree	(5) Owning Home	(6) Internet Home
A. Test for Student Assignment						
Kolmogorov-Smirnov test (No. failed / total tests)	7/229	18/343	0/343	0/341	9/349	0/336
χ^2 goodness of fit test (No. failed / total tests)	9/229	16/343	9/343	14/341	15/349	8/336
B. Test for Teacher Assignment						
Female Teacher	0.018 (0.013)	-0.032 (0.018)	0.004 (0.019)	0.033 (0.023)	-0.016 (0.016)	-0.010 (0.015)
Teacher Experience	0.000 (0.001)	-0.001 (0.001)	0.002 (0.001)	0.000 (0.001)	0.001 (0.001)	-0.001 (0.001)
Graduate Degree	-0.014 (0.020)	0.014 (0.017)	0.029 (0.019)	0.045* (0.019)	0.028 (0.019)	0.026 (0.017)
Teacher's College Graduate	-0.019 (0.020)	-0.012 (0.014)	0.000 (0.014)	0.008 (0.014)	0.003 (0.016)	-0.013 (0.015)
Teacher's Union Member	-0.008 (0.026)	0.014 (0.018)	-0.012 (0.023)	-0.000 (0.025)	-0.000 (0.017)	0.018 (0.019)
N	530	856	848	720	909	671
R^2	0.210	0.184	0.064	0.221	0.053	0.277

Notes: Each column represents a separate regression. The dependent variable is the empirical p-value from the resampling described in the text, for student characteristics, which are male student, having married parents, having father with BA degree or higher, having mother with BA degree or higher, owning their home, and having access to Internet home in Columns 1 through 6, respectively. Independent variables are teacher gender, teacher experience measured in years, dummies for graduate degree, teachers' college graduate, and teachers union member. Each specification controls for subject and school fixed effects. Standard errors in parentheses are clustered at school level. The Kolmogorov-Smirnov and χ^2 goodness of fit test results indicate the number of tests of the uniformity of the distribution of p-values that failed at the 5 percent level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.1.2 Pearson's χ^2 Tests

We next turn to testing random assignment with respect to observable characteristics by conducting a series of Pearson's χ^2 tests for independence of a variety of characteristics and the classroom to which they are assigned. Tested characteristics include student's gender, parents' marital status, parents' education, as well as whether parents own their own home and whether student's home has access to the Internet, as proxies for family resources. Parents' education has seven categories and the other variables are indicator variables.

<center>(a) All Subjects   (b) Korean Language</center>

<center>(c) English Language   (d) Mathematics</center>
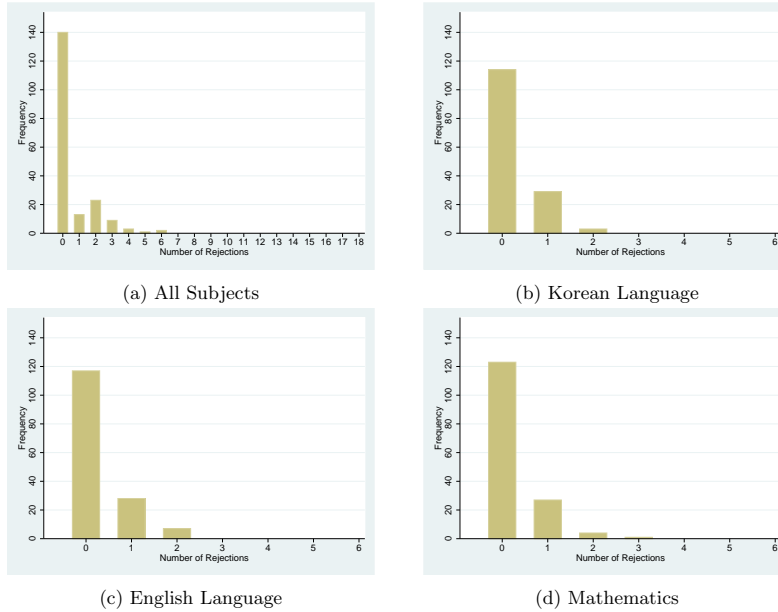
<center>Figure 1: Rejections of Classroom-Characteristic Independence by School</center>

We perform 2,082 Pearson's χ^2 tests across six characteristics and 453 school-subject combinations.[11] We find that 208 (9.99%), 115 (5.5%), and 38 (1.8%) of these p-values are lower than or equal to 10%, 5%, and 1%, respectively. This provides further evidence for the random assignment mechanism described in Section 2.1.

To check whether the rejections are concentrated in particular schools, we examine distributions of the number of rejections by school. Figure 1 shows the distributions for all subjects and each subject. Two schools have a total of six rejections in all subjects combined and one school has five rejections. Only one school has as many as three rejections in one subject, suggesting that no schools that are failing to comply with the random assignment of students to classrooms. Further, omitting the three schools with five or six total rejections from our estimates does not affect the results.

3.1.3 Difference in Mean Characteristics

Another approach is to compare the groups of students taught by same- and opposite-gender teachers. If the students are randomly assigned to the teachers of the same and opposite gender, then the two groups should look similar in terms of observable characteristics.

Table 2 presents sample means from our data, with each observation as a

---

[11]Some combinations cannot be tested due to missing variables, only a single classroom from the school remaining in the sample, or the school itself being single-sex.

student-teacher pair. Recall that the randomization in our sample is *within* schools, though even when looking across schools, the results are fairly well-balanced. In Panel A, the characteristics for female and male students are presented separately by teacher gender, demonstrating that students are not more likely to be assigned to a teacher of the same gender based on observable characteristics. For male students, there is a statistically significant difference for home ownership, but it is economically small. Moreover, since random assignment was done within schools, adjusting for school fixed effects eliminates the significance of this difference. We also show the mean standardized test scores by group as a preview of our results. Female students perform substantially better than male students overall, but particularly when they have female teachers. Meanwhile, male students are not greatly affected by the gender of their teacher. In Panel B, we compare teachers' characteristics when assigned female and male students. As in most schools around the world, female teachers are much more prevalent in our sample, but there are no significant differences in the types of teachers assigned to students of different gender. These results further show that students and teachers are randomly assigned to classrooms irrespective of gender matches.[12]

---

[12]We also follow Carrell, Page and West (2010) in examining whether the student characteristics in Table 1 predict teacher gender; they are jointly insignificant at $p = 0.31$.

Table 2: *Comparison of Mean Characteristics*

A. Student Characteristics

	Female Students				Male Students			
	Female Teachers	Male Teachers	P-value	Observations	Female Teachers	Male Teachers	P-value	Observations
Married Parent	0.903 (0.007)	0.899 (0.006)	0.673	10,326	0.901 (0.006)	0.893 (0.007)	0.315	10,059
Dad w/ College or More	0.263 (0.022)	0.223 (0.023)	0.086	10,073	0.269 (0.021)	0.223 (0.025)	0.115	9,751
Mom w/ College or More	0.143 (0.017)	0.126 (0.018)	0.338	10,135	0.145 (0.016)	0.119 (0.020)	0.270	9,688
Parents Own Home	0.711 (0.015)	0.732 (0.016)	0.264	10,476	0.716 (0.013)	0.748 (0.013)	0.033	10,202
Internet Access at Home	0.916 (0.006)	0.912 (0.007)	0.616	10,272	0.907 (0.006)	0.903 (0.007)	0.580	9,939
Standardized Test Score	0.114 (0.030)	0.057 (0.044)	0.181	11,925	-0.100 (0.037)	-0.089 (0.042)	0.827	12,306

B. Teacher Characteristics

	Female Teachers				Male Teachers			
	Female Students	Male Students	P-value	Observations	Female Students	Male Students	P-value	Observations
Teacher Age	36.2 (0.575)	35.8 (0.567)	0.515	15,719	43.0 (0.963)	43.6 (0.811)	0.569	8,406
Teacher Experience (year)	11.8 (0.605)	11.4 (0.614)	0.524	15,265	17.0 (0.975)	18.0 (0.952)	0.414	8,315
Teachers College Graduate	0.759 (0.030)	0.736 (0.030)	0.467	15,794	0.633 (0.063)	0.614 (0.048)	0.777	8,406
Graduate Degree Teacher	0.207 (0.029)	0.204 (0.032)	0.923	15,794	0.454 (0.052)	0.401 (0.057)	0.382	8,406
Homeroom Teacher	0.787 (0.030)	0.824 (0.024)	0.216	15,672	0.601 (0.049)	0.660 (0.048)	0.318	8,319
Regular Full Time Teacher	0.956 (0.013)	0.962 (0.011)	0.518	15,647	0.967 (0.022)	0.980 (0.015)	0.633	8,360

Notes: Each p-value is for a test of equality of means. Standard errors in parentheses are clustered at the school level.

3.2 Specifications

To analyze the effect of teacher-student gender interaction, we estimate the following linear regression equation:

$$y_{ijsb} = \beta_0 + \beta_1 fs_i + \beta_2 ft_j + \beta_3 fs_i ft_j + X_{ij}\gamma' + \alpha_s + \alpha_b + \varepsilon_{ijsb}, \qquad (1)$$

where y_{ijsb} is the test score of student i who was taught by teacher j in school s for subject b. The test scores are normalized in each subject to have mean zero and variance of one. Because the scores in Korean language, English language and math are pooled together, we also include subject fixed effects α_b. fs_i and ft_j are indicator variables having value of one when student i and teacher j, respectively, are female. X_{ij} is a vector of student and teacher characteristics including indicators for married parents and parental education; teacher characteristics include indicators for graduate degree and graduation from a teachers college, and indicators for teacher experience of two years or below, two to three years, three to four years, four to five years, and five years or more. α_s are school fixed effects, included since random assignment of students is done within schools.

We estimate the equations by ordinary least squares (OLS), which produces unbiased estimates given the random assignment of students and teachers to classrooms. Standard errors are clustered at the school level to accommodate correlations among students within the same schools; we obtain similar standard errors clustering at the classroom level or with two-way clustering at the student and teacher level.

β_1 is the average difference in academic achievement for female compared to male students with male teachers, while β_2 indicates the impact of a female versus male teacher on performance for male students. The total effect of having a female teacher for female students can be obtained by adding β_2 to β_3, with β_3 as the differential effect on female students, as compared to male students, of having a female teacher. This last coefficient is the change in the gender gap between female and male students when switching from a male teacher to a female teacher.

4 Results

4.1 Main Effects

Table 3 presents the coefficients from estimating variations of Equation (1). We begin in Column 1, with school and subject fixed effects. The coefficient on the female student variable indicates that female students perform better than male student by about 0.15 of a standard deviation on average across Korean language, English language, and math when paired with a male teacher. The change in the performance gender gap between females and males when switching from a male teacher to a female teacher, as indicated by the interaction effect between female student and female teacher, is 0.098 standard deviations.

This total effect is comprised of a small and statistically insignificant decrease in male performance of 0.021 standard deviations and an increase in female performance of 0.076. This widening of the gender gap is substantial, representing more than a third of a year of schooling based on the general rule of thumb that 1% of a standard deviation of performance is roughly equivalent to 10 days of schooling (Carlsson et al., 2015).

Table 3: *Main Results*

	Single Teacher Only						Single and Multiple Teachers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Student	0.147^{***}	0.144^{***}	0.157^{***}				0.148^{***}
	(0.030)	(0.030)	(0.033)				(0.029)
Female Teacher	-0.021	-0.022	-0.052	-0.025			-0.028
	(0.029)	(0.029)	(0.052)	(0.094)			(0.025)
Female Student × Female Teacher	0.098^{***}	0.105^{***}	0.086^{**}	0.104^{**}	0.093^{*}	0.094^{*}	0.103^{***}
	(0.032)	(0.032)	(0.035)	(0.051)	(0.049)	(0.049)	(0.030)
Constant	-0.093^{***}	-0.110^{***}	-0.047	-0.107	-0.033^{**}	-0.033^{**}	-0.087^{***}
	(0.021)	(0.040)	(0.100)	(0.111)	(0.016)	(0.016)	(0.019)
Observations	24,231	23,580	23,580	23,580	24,231	24,231	30,673
R^2	0.110	0.112	0.123	0.862	0.862	0.859	0.106
School FEs	Yes	Yes					Yes
Subject FEs	Yes	Yes					Yes
Teacher Ctrl		Yes	Yes	Yes			
Sch-Sbj FEs		Yes	Yes	Yes	Yes	Yes	
Student FEs			Yes	Yes	Yes	Yes	
Cls-Sbj FEs					Yes		
Teacher FEs						Yes	

Notes: Each column represents a separate regression. Columns 1 through 6 are for students taught by a single subject teacher, with each variable as a binary indicator. Column 7 includes students taught by either single or multiple subject teachers, with the female teacher variable representing the fraction of the student's subject teachers who are female. Student and teacher level variables are omitted in Columns 4 through 6 because of collinearity. Standard errors in parentheses are clustered at school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Including teacher background controls in Column 2 does not change the coefficients of our interest much.[13] We replace school and subject effects with school-by-subject fixed effects beginning with Column 3. In Column 4, we add student fixed effects to test for the presence of unobserved student characteristics correlated with the variables of interest. These also subsume classroom fixed effects and also control for peer effects, since students do not change classrooms. Their inclusion does not change the gender gap appreciably. In Column 5, we follow Fairlie et al. (2014) and include classroom-by-subject fixed effects to account for the possibility of subject-specific classroom shocks. The results are unchanged. Finally, in Column 6, we add teacher fixed effects to test whether unobserved teacher characteristics are driving our results, despite random assignment. The teacher-student gender interaction coefficient remains the same size and is statistically significant at $p = 0.056$. Taken together with the evidence in Section 3.1, the stability of this coefficient strongly suggests that the random assignment to classrooms in South Korea is, indeed, in place. As such,

---

[13]Results are similar when including a variety of student background characteristics, but survey nonresponse reduces the sample substantially.

the interpretation of our results is free of the potential problems caused by sorting on unobservable characteristics.

Our findings are comparable in magnitude to those in Dee (2007) and Carrell et al. (2010). Dee's estimate of the increase in the gap between female and male students when assigned to a female teacher is about 0.092, with opposing positive and negative effects of similar size for female and male students, respectively. While our effect is concentrated on improvements for female students, it is quite similar in magnitude for a similar length of exposure to that year's teachers (about one semester). Carrell et al.'s effect, for somewhat less than one semester of exposure to a female professor, is 0.097 standard deviations with a reduction in male performance of 0.050 standard deviations.

To show that our results are not affected by the 6,442 observations that were dropped due to students having multiple subject teachers, we include them in Column 7. This specification corresponds to that in Column 1, but the female teacher variable represents the fraction of the student's subject teachers who are female. About 90% of these additional observations are groups of two teachers; nearly all of the remaining ones have three teachers. The results are essentially unchanged, with the gender gap increasing by 0.10 standard deviations when all of a female student's teachers are female themselves.[14]

4.2   Effects by Subject

The gender gap differs by subject, with female students generally performing substantially better than males in language arts but about even or slightly worse in science and mathematics (OECD, 2015). Teachers' impacts may be greater in mathematics, given negative stereotypes about female mathematical ability; for example, Spencer, Steele and Quinn (1999)'s experimental study shows that negative stereotypes regarding the mathematical ability of female students negatively affects their test scores.

To test whether our results vary by subject, we fully interact the specification in Column 1 of Table 3 with indicators for English and mathematics. The coefficients, in Panel A of Table 4, show the full set of interactions. We note that female students perform far better than male students in Korean (0.34 standard deviations) and English (0.20 standard deviations), and about evenly in math (-0.04 standard deviations), with the last of these differences being statistically insignificant. In Panel B, we combine the relevant coefficients to calculate the change in the gender gap between female and male performance when switching from a male to female teacher. For Korean language courses, the gender gap between girls and boys does not widen significantly, though it does

---

[14]We also estimate the specification in Column (1) excluding 7,964 student-teacher observations at single-sex schools, and an additional 1,620 observations assigned to single-sex classrooms in coeducational schools. The female teacher-female student coefficient for the former sample is 0.087 (s.e. = 0.035) and 0.083 (s.e. = 0.035) for the latter. We also test whether our results differ for students in rural and urban areas and by parental education; no consistent patterns emerge and none of the differences in the interaction variable are statistically significant.

Table 4: Results by Subject

A. Main & Interaction Effects

Female Student	0.336***
	(0.040)
Female Teacher	−0.035
	(0.039)
Female Student × Female Teacher	0.042
	(0.047)
English × Female Student	−0.135***
	(0.048)
Math × Female Student	−0.380***
	(0.046)
English × Female Teacher	0.049
	(0.052)
Math × Female Teacher	0.021
	(0.048)
English × Female Student × Female Teacher	0.057
	(0.060)
Math × Female Student × Female Teacher	0.042
	(0.061)
Constant	−0.156***
	(0.027)

B. Change in the Performance Gap

Korean Language	0.042
	(0.047)
English Language	0.099**
	(0.045)
Math	0.084**
	(0.041)
Observations	24,231
R^2	0.115

Notes: Estimates include English and math fixed effects as well as school fixed effects. Standard errors in parentheses are clustered at school level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

for English and math; however, there are no statistically significant differences between these effects.

5   Evidence on Mechanisms

To investigate the mechanisms underlying the positive impact of female teachers on female students, we examine a series of student responses about classroom interactions, as well as questions about private tutoring asked of parents. There are numerous such questions in the KEDI data, but we chose to focus on those that may distinguish student- and teacher-centered mechanisms. The results in Table 5 correspond to the specification in Column 1 of Table 3. In Columns 1-4, the dependent variable is an indicator for whether the student agrees or agrees strongly with the following sentiments, respectively: the teacher provides students with equal opportunity to participate in class; the teacher encourages students to express themselves; I feel comfortable asking the teacher a question; and I ask many questions in this class. The first two questions are proxies for teacher-centered mechanisms – that is, they are about the teacher's behavior. The next two questions are about the student's behavior, as are the estimates in Columns 5-7. In Column 5, the dependent variable is a continuous measure of hours of study in that subject (excluding hours spent at tutoring). Column 6, asked of parents, reports the likelihood of receiving tutoring in the subject; note that over 60% of students receive tutoring. In Column 7, we examine the effect on the log of tutoring expenditures, conditional on reporting any. This variable, reported by parents as well, provides an indication of tutoring intensity, both in terms of time and personal attention. Finally, Column 8 is the impact on student's self-report that the subject is his or her favorite. This can be influenced by both student- and teacher-centered mechanisms, and provides a useful proxy for the student's overall response to the teacher.

Female students are significantly less likely to feel as if they have an equal opportunity to participate or are encouraged with male teachers, but this negative outcome is eliminated when the teacher is female. On the other hand, while all students report greater comfort in asking questions when the teacher is female, there is no additional effect on female students; they also are somewhat less likely to report asking many questions. There is no effect on hours of study, nor on either tutoring outcome variable. Overall, female students are significantly more likely to report that the subject is their favorite when the teacher is female.[15]

Finally, we examine whether the effects differ by the proportion of female students in the classroom. A greater number of female students means that a female teacher can give less attention to each individual student; on the other hand, a higher proportion of female students may enable the teacher to provide a more welcoming environment for girls. We begin by examining whether the

---

[15]We estimated versions of Table 5 with interactions by subject, as in Table 4. While effects tended to be larger in math and, to a lesser extent, English classes relative to Korean classes, none of the differences were statistically significant.

impacts of the teacher-student gender match are greater at single-sex schools. While the interaction effect is somewhat larger (0.037 standard deviations), it is not statistically significant ($p = 0.50$); similar results are obtained when comparing single-sex to coeducational classrooms. Female students in classrooms with above-median numbers of females do perform better with female teachers (0.036 standard deviations), but once again the difference is not significant ($p = 0.57$). We also interact the proportion of female students in a coeducational classroom with the teacher-student gender interaction. Once again, the difference is fairly large but statistically insignificant. For example, a ten percentage point increase in female representation increases the interaction effect by 0.013 standard deviations, with a standard error of 0.023 standard deviations. Without making too much of these differences, they suggest that female teachers may be conducting their classrooms differently in a manner that has a positive impact on female students.

Taken together, the results on female students' responses to female teachers and the somewhat-large effects for classrooms with more female students provide suggestive evidence that the increase in female student performance with female teachers is driven by teacher rather than student behavior.

6 Conclusion

Understanding the effect of teacher-student gender interactions on student's academic achievement is important not only for evaluating policies to close the gender gap in academic achievement, but also to enhance understanding of the education production function. However, it is difficult to estimate a student-teacher gender match effect free of selection bias because of the nonrandom sorting of students.

In this study, we estimate the impact of teacher-student gender matches on academic achievement using the random assignment of students in South Korea. We find that the performance gender gap between female and male students increases dramatically when switching from male to female teachers (0.098 standard deviations). Male students do not appear to benefit appreciably from a teacher of the same gender, but female students' performance increases by about 8% of a standard deviation when they are taught by a female teacher. This effect is large, and driven primarily by performance in English and mathematics courses. We also provide evidence that teacher behavior drives this increase in student achievement.

Our findings are consistent with the results of Dee (2007) and Carrell et al. (2010). Combining these similarities, the random assignment nature of our approach, and the evidence on South Korea's attitudes towards gender equality (Brandt, 2011), we conclude that these interactions reflect genuine changes in the classroom environment that are not necessarily driven by the environment being studied.

Table 5: *Effects on Student and Teacher Behavior*

	Equal Chance to Participate	Teacher Encourages Expression	Comfort Asking Questions	Asks Many Questions	Hours of Study	Receives Tutoring	Log Tutoring Expenditures	Favorite Subject
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Student	−0.050***	−0.056***	−0.029	−0.020**	−0.137	−0.079***	0.031	−0.007
	(0.018)	(0.017)	(0.019)	(0.009)	(0.094)	(0.018)	(0.048)	(0.016)
Female Teacher	0.039*	0.012	0.050**	0.029***	−0.046	−0.014	−0.039	−0.005
	(0.022)	(0.021)	(0.021)	(0.010)	(0.061)	(0.014)	(0.031)	(0.015)
Female Student × Female Teacher	0.060***	0.063***	0.008	−0.021*	0.063	0.017	−0.080	0.041**
	(0.022)	(0.020)	(0.023)	(0.011)	(0.097)	(0.018)	(0.049)	(0.020)
Constant	0.379***	0.332***	0.314***	0.099***	1.857***	0.623***	2.223***	0.286***
	(0.015)	(0.014)	(0.014)	(0.007)	(0.052)	(0.012)	(0.027)	(0.011)
Observations	23,773	23,737	23,755	24,065	24,227	17,812	6,788	23,900
R^2	0.053	0.055	0.036	0.028	0.044	0.120	0.194	0.028

Notes: Each column represents a separate regression and includes subject and school fixed effects. The response variables for Columns 1 through 4, and 8 are indicators taking value of one if a student agrees or strongly agrees with the statement that the subject teacher gives all students an equal opportunity to participate in class; the subject teacher encourages students to be creative and express themselves; I feel comfortable asking the subject teacher a question when the lecture is difficult to understand; I ask many questions in this class; this subject is one of my favorites. The outcome variable in Column 5 is self-reported study hours per week for the subject, excluding hours spent at tutoring. The outcome in Column 6 is an indicator for receiving tutoring and that for Column 7 is the log of tutoring expenditures. Column 7 is regressed conditional on positive expenditures. Standard errors in parentheses are clustered at school level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik, "The Effect of Teacher Gender on Student Achievement in Primary School," *Journal of Labor Economics*, 2015, *33* (1), 63–89.

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

Brandt, M. J., "Sexism and gender inequality across 57 societies," *Psychological science*, 2011, *22* (11), 1413–8.

Byun, Sooyong and Kyungkeun Kim, "Educational inequality in South Korea: The widening socioeconomic gap in student achievement," *Research in Sociology of Education*, 2010, *17* (155-182).

Canes, Brandice J. and Harvey S. Rosen, "Following in Her Footsteps? Faculty Gender Composition and Women's Choices of College Majors," *Industrial and Labor Relations Review*, 1995, *48* (3), 486–504.

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth, "The effect of schooling on cognitive skills," *Review of Economics and Statistics*, 2015, *97* (3), 533–547.

Carrell, Scott E. and James E. West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 2010, *118* (3), 409–432.

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

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

Cho, Insook, "The effect of teacher student gender matching: Evidence from OECD countries," *Economics of Education Review*, 2012, *31* (3), 54–67.

Dee, Thomas S., "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, 2007, *42* (3), 528–554.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos, "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom," *American Economic Review*, 2014, *104* (8), 2567–91.

Good, Phillip I., *Resampling Methods: A Practical Guide to Data Analysis*, 3rd ed. Boston: Birkhäuser, 2006.

Hoffmann, Florian and Philip Oreopoulos, "A professor like me: the influence of instructor gender on college achievement," *Journal of Human Resources*, 2009, *44* (2), 479–494.

Holmlund, Helena and Krister Sund, "Is the gender gap in school performance affected by the sex of the teacher?," *Labour Economics*, 2008, *15* (1), 37–53.

Jackson, C Kirabo, "Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago," *Journal of Public Economics*, 2012, *96* (1), 173–187.

Jones, Susanne M. and Kathryn Dindia, "A meta-analytic perspective on sex equity in the classroom," *Review of Educational Research*, 2004, *74* (4), 443–471.

Kang, Changhui, "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea," *Journal of Urban Economics*, 2007, *61* (3), 458–495.

Korea Legislation Research Institute, "Enforcement Decree of the Elementary and Secondary Eduation Act, Article 68 (Methods for Entering Middle Schools)," *Presidential Decree No.22712*, Mar 18, 2011.

Lavy, Victor and Edith Sand, "On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers Stereotypical Biases," *NBER Working Paper No. 20909*, 2015.

Lee, Soohyung, Lesley J Turner, Seokjin Woo, and Kyunghee Kim, "All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement," *NBER Working Paper No. 20722*, 2014.

Lehmann, Erich L. and Joseph P. Romano, *Testing Statistical Hypotheses (Springer Texts in Statistics)*, 3rd ed., Springer, April 2005.

Muralidharan, Karthik and Ketki Sheth, "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers," *Journal of Human Resources*, 2016, *51* (2), 269–297.

OECD, *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing, 2015. http://dx.doi.org/10.1787/9789264229945-en.

Paredes, Valentina, "A teacher like me or a student like me? Role model versus teacher bias effect," *Economics of Education Review*, 2014, *39* (2), 38–49.

Park, Hyunjoon, Jere R Behrman, and Jaesung Choi, "Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in Seoul high schools," *Demography*, 2013, *50* (2), 447–469.

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