

# Persistent Effects of Teacher-Student Gender Matches

Jaegeum Lim\*  
Korean National Assembly

Jonathan Meer  
Texas A&M University  
and NBER

August 23, 2017

## Abstract

We exploit data from middle schools in Seoul, South Korea, where students and teachers are randomly assigned to classrooms, and find that female students taught by a female versus a male teacher score higher on standardized tests compared to male students even four years later. We also find that having a female math teacher in 7th grade increases the likelihood that female students take higher-level math courses, aspire to a STEM degree, and attend a STEM-focused high school. These effects are driven by changes in students' attitudes and choices.

## 1 Introduction

While women have made great strides in fields previously dominated by men, some gender gaps – especially in STEM fields – remain. Lack of female representation can have compounding effects, in which a dearth of role models or general attitudes and expectations discourage female students from entering these fields, making it difficult to increase participation (Leslie et al., 2015).

One possible source of these discrepancies may be the interaction of student and teacher gender, in which female students are treated differently by or react differently to female and male teachers. For example, Spencer, Steele and Quinn (1999) show that negative stereotypes of girls' math ability undermine girls' performance on math exams. Most research has focused on contemporaneous effects, namely, how students' current-year outcomes are affected by their current teacher(s). The results are mixed: many studies find positive and sizable impacts of female teachers on female students (see, for example, Muralidharan and Sheth (2016), Dee (2007), and Lim and Meer (forthcoming)), though others do not (Antecol, Eren and Ozbeklik (2015), Holmlund and Sund (2008), and Cho (2012)). Yet the impact on current outcomes is of limited relevance if the

---

\*Author emails are [jlim@na.go.kr](mailto:jlim@na.go.kr) and [jmeer@tamu.edu](mailto:jmeer@tamu.edu).

effects fade over time. Evidence on longer-run impacts is limited, though indicative that effects can be substantial; [Lavy and Sand \(2015\)](#) find that primary school teachers' bias towards boys can have impacts through high school. At the higher education level, [Carrell, Page and West \(2010\)](#) find that female cadets at the United States Air Force Academy who are randomly assigned to female instructors for introductory science and math courses are more likely to persist in those fields.

The main issue plaguing identification of the impact of student-teacher gender matches is that, in most contexts, students are not randomly assigned to teachers. This leads to potential confounds in which, for example, female teachers may be systematically assigned to lower-achieving students, or female students who are more likely to succeed with female teachers are assigned to those classrooms. In these cases, any teacher-student gender match effects may reflect sorting rather than, for example, role-model effects. While several of the studies listed above used environments featuring random assignment, there remains little evidence on whether the teacher-student gender match has a long-run impact when taking place at more formative ages – after all, it may be too late to substantially affect field of study by the time a student arrives in college. Further, there is little evidence on the mechanisms through which these effects are operating.

In this paper, we study the longer term effects of teacher-student gender matches at the secondary level. We avoid the problem of nonrandom sorting using a unique Korean middle school practice: the random assignment of students into a classroom each year. Our data track students from 7th to 11th grades and enable us to investigate how these effects change over time, as well as how they affect important outcomes like preparation for postsecondary study.

We find that the presence of a female teacher substantially increases female students' test scores compared to male students, and this effect persists at least through 11th grade. Our long-lasting gender gap effects are somewhat surprising, since the effects of educational interventions generally fade out fairly quickly ([Jacob, Lefgren and Sims, 2010](#)). We show that the mechanism behind these persistent effects appear to be greater focus and participation by female students, as well as selection into higher-quality high schools.

## 2 Data

### 2.1 Institutional Background

Several features of the South Korean educational system make it ideally suited to study the impact of teacher-student gender matches. First, elementary school graduates are assigned to a local middle school (spanning 7th to 9th grades) by lottery.<sup>1</sup> Second, middle school students are randomly assigned to a physical

---

<sup>1</sup>While students in some districts can submit a list of schools among which they will be randomly assigned, students in Seoul are not permitted to do so ([Korea Legislation Research Institute, 2011](#)).

homeroom classroom, through which subject teachers rotate to give lessons. Due to strong social norms and government policies, the goal is to produce homogeneous homeroom classrooms in terms of academic ability (Kang, 2007). The most common approach is to order students by the previous year’s academic performance, with the leading student to the first classroom, the second-ranked student to the second classroom, and so on, and schools report adhering to these expectations (Lim and Meer, forthcoming).

Subject teachers are assigned to classrooms in a manner unrelated to the characteristics of either the teachers or the classroom’s students. In general, the physical classrooms are split between the subject teachers by, for example, dividing odd- and even-numbered rooms; teacher characteristics are not correlated with student characteristics (Lim and Meer, forthcoming). This quasi-random classroom and teacher assignments produce the random variation in teacher student matching within a school.

Over the past decade, there has been increasing use of ability tracking in South Korea (Byun and Kim, 2010). In our data, ability grouping is common for math and English, but infrequent for Korean. In the presence of ability grouping, the students move to the classroom with their ability group for that subject before returning to their original homeroom classroom. Figure 1 gives an example, with high ability students in Classroom 2 moving to Classroom 1 and low ability students in Classroom 1 moving to Classroom 2 for this subject. This might threaten our identification strategy, which rests on the random assignment of students to teachers. Including school-by-subject fixed effects might be insufficient to account for this potential source of bias. We therefore estimate specifications with school-by-subject-by-ability-group fixed effects, enabling us to compare similar sets of students. In practice, our results are unaffected by the inclusion of finer fixed effects, suggesting that ability grouping does not lead to sorting correlated with teacher-student gender matches and does not bias our results. Our identification thus remains driven by the random assignment of students and teachers to classrooms.

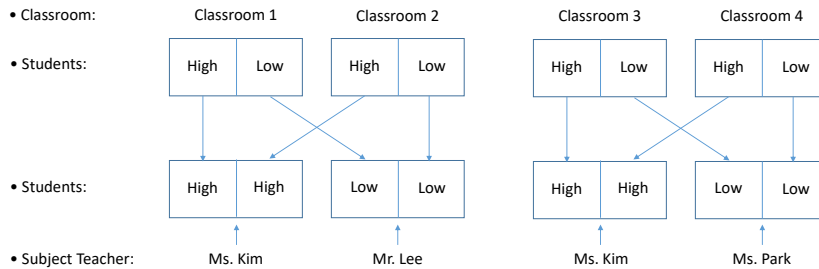


Figure 1: Example of Ability Group Formation

While compulsory education ends in ninth grade, nearly all students in South Korea go on to high school. There are two rounds of admissions to high school in

Seoul. Admission is determined through an application process in the first round (Seoul Metropolitan Office of Education, 2012). Students can apply to one of 35 magnet and private autonomous high schools, six art high schools, one athletic high school, and 74 vocational high schools. Schools select students based on their academic performance and recommendations from principals and teachers. If selected, students must attend the school to which they were admitted.

The second round consists of a lottery for 19 autonomous public high schools, 19 science-focused high schools, 4 art-focused high schools, and 183 general academic high schools. Students can list preferences for one autonomous school, one science- or art-focused school, and up to four general schools. Students receive some preference for schools within their administrative district. The lottery clears entry into the autonomous public and specialized schools first, and then into the general academic high schools. Students are guaranteed entry into a school. Conditional on entering 10th grade, 95.2% of students in 2011 graduate; only 0.32% repeat a grade.

One year after high school entry, students choose among academic tracks, irrespective of the type of school. Most schools provide a math-science track and humanities-social science track, excepting specialized science and foreign-language schools. Students focus on advanced courses within a given track. Students are free to choose their academic track; past test scores, student characteristics, and so on are not taken into account by the school in restricting a student's choice. Changing tracks is rare; students apply to postsecondary institutions the following year, in twelfth grade, and it is difficult to catch up on the requisite coursework. Notably, applying to a STEM department at a postsecondary institution requires scores in advanced math and science courses. Therefore, if teacher-student gender matches in middle school – a particularly formative time period (Berenbaum, Martin and Ruble, 2008) – have persistent effects into high school, they are likely to have lifelong impacts on academic and career choices.

## 2.2 Data Set

Our data set is the Seoul Education Longitudinal Study of 2010 (SELS2010), which surveyed 7th grade students and their teachers in 2010; data through 11th grade are available. Subject teachers in math, English language, and Korean language are linked to the students in their seventh through tenth grades.<sup>2</sup>

Students in the seventh grade panel were sampled by stratified two-stage cluster sample design; first, 74 middle schools were randomly chosen from the population of 370 public or private middle schools in Seoul, excluding two middle schools that are operated by the central government and one athletic middle school. Sixty-two of the sampled schools are coeducational, seven are all-boys, and five are all-girls schools. Two classrooms were then drawn randomly within the sample school. 4,544 of the 5,065 targeted students responded to the survey

---

<sup>2</sup>Errors in data collection in eleventh grade mean that those links are not available in that year.

in 2010. Attrition reduced the sample to 4,347, 4,162, 3,541, and 3,394 from 2011 through 2014, respectively; we show later that this attrition does not drive our results. Students in the sample advanced to high school in 2013, with 3,017 going to academic and 524 to vocational high schools. 2,893 and 501 students in academic and vocational high schools remained in the sample through 2014.

Our primary unit of observation is a student-teacher pairing, with multiple observations per student, allowing us to use within-student variation in teacher gender in a given year. Of 13,632 possible student-teacher matches in seventh grade, 3,364 observations cannot be linked due to teacher non-response. An additional 74 observations have missing values in test scores, student gender, and teacher gender. We compare remaining 10,194 observations, representing 4,146 students and 488 teachers, with those that are dropped and find no significant differences between them in students' predetermined characteristics. 1,885 (45.5 percent) students and 402 (82.4 percent) teachers are female. 11 percent of these teacher-student pairs are male-male, 8 percent are male-female, 44 percent are female-male, and 37 percent are female-female.

Student data include standardized test scores for each subject (taking place after the first semester of the school year); we normalize these to have mean of zero and standard deviation of one for ease of interpretation. We also have data on student background characteristics, course-taking in high schools, and track choice, as well as on teacher characteristics including experience and education.

## 3 Empirical Approach

### 3.1 Tests of the Identifying Assumptions

While the institutional characteristics of our setting provide the random variation in teacher-student gender matches necessary for unbiased estimation, it is instructive to examine empirical evidence to validate the identification strategy.

First, we test differences in the characteristics of student-teacher pairings when the teacher is female or male. Panel A of Table 1 presents these results and show no meaningful differences. Panel B compares teacher characteristics by student gender; there are no significant differences between the sets of observations. As another way to examine whether there are systematic differences in how teachers are assigned to students, we regress teacher gender on students' observable characteristics, controlling for school-by-subject-by-ability-group fixed effects. The results are shown in Table 2. Teacher gender is not correlated with students' predetermined characteristics, consistent with random assignment of students to teachers.

Next, we show that that subject teachers' characteristics are not correlated with the students' ability and that students are not assigned to physical home-room classrooms on the basis of their ability. We conduct this exercise for ninth graders in order to use test scores from previous years as a proxy for ability, thus demonstrating that students are not systematically assigned to classrooms

Table 1: Comparison of Mean Characteristics

<i>A. Student Characteristics</i>				
	w/ FT	w/ MT	P-Value	Observations
Female Student	0.46 (0.02)	0.42 (0.05)	0.37	10,258
Married Parents	0.87 (0.01)	0.87 (0.01)	0.85	10,029
Both Parents Work	0.47 (0.01)	0.46 (0.01)	0.20	9,984
Family Income	489.5 (18.7)	480.8 (19.5)	0.65	9,122
Parents w/ B.A. or Higher	0.57 (0.03)	0.58 (0.03)	0.67	9,427
Number of Siblings (Including Students)	1.85 (0.01)	1.84 (0.02)	0.47	10,005
<i>B. Teacher Characteristics</i>				
	w/ FS	w/ MS	P-Value	Observations
Female Teacher	0.83 (0.02)	0.80 (0.03)	0.38	10,258
Graduate Degree	0.33 (0.03)	0.33 (0.03)	0.90	10,258
Teacher's College	0.63 (0.03)	0.62 (0.03)	0.60	10,172
Admin Teacher	0.15 (0.02)	0.19 (0.02)	0.03	10,110
Over Mid Age	0.53 (0.03)	0.55 (0.03)	0.48	10,170
Teaching less than 5 Yrs	0.26 (0.03)	0.26 (0.03)	0.79	9,463

Each p-value is for a test of equality of means. Standard errors in parentheses are clustered at the school level. Family income is measured in 10,000 KRW, approximately 865 USD in 2010.

with male or female teachers based on their previous performance.<sup>3</sup>

We use resampling techniques to examine whether actual classroom assignments appear unusual relative to artificially-constructed classrooms (for more details, see [Lehmann and Romano \(2005\)](#), [Good \(2006\)](#), and [Carrell and West \(2010\)](#)). We randomly draw, without replacement, 10,000 synthetic classrooms from the sample of all students within the school. For each of these synthetic classrooms, we compute total test score of its students. We calculate an empirical p-value measuring the proportion of artificial classrooms with a value greater than the actual classroom in the data. We conduct this exercise using test scores from all three subjects. Under random assignment, the empirical p-values will be uniformly distributed, since any p-value will be observed with equal probability. We test whether the distribution of the empirical p-values is uniform with Kolmogorov-Smirnov and  $\chi^2$  goodness of fit tests. The results, in

<sup>3</sup>Data on the assignment to *physical* homeroom classrooms was not available for 7th and 8th grades, but subject teacher-student linkages are present. Given the nature of random assignment in South Korea, physical classroom assignments are not necessary for our identification strategy.

Table 2: Likelihood of Having a Female Teacher

	Coefficient	S.E.
Female Student	0.012	(0.009)
Married Parents	0.003	(0.007)
Family Income	0.000	(0.000)
Dad w/ B.A. or Higher	-0.003	(0.004)
Mom w/ B.A. or Higher	-0.002	(0.004)
Attended Prv. Ele. School	0.011	(0.010)
Number of Siblings	0.004	(0.003)
Observations	8,426	
$R^2$	0.798	

The dependent variable is an indicator for having a female teacher in 7th grade, regressed on student characteristics, controlling for school by subject by ability group level fixed effects. Standard errors in parentheses are clustered at school level.

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

Table 3A, show that reject none of 222 school-by-subject p-value distributions deviate significantly from the uniform distribution at the 5 percent level using the Kolmogorov-Smirnov test; 6 (2.7 percent) do so using the  $\chi^2$  goodness of fit test. We do not find evidence of nonrandom assignment of students into classrooms by academic ability.

We also check the random assignment of teachers into classrooms with respect to student's academic ability; for each of three test scores, we regress the empirical p-values on average characteristics of teachers visiting the classroom, controlling for school fixed effects to accommodate the random assignment within a school. As is shown in Table 3B, none of 24 coefficients is statistically significant at the 5 percent level. Therefore, there is little evidence of non-random assignment of teachers into classrooms in terms of student's academic ability.

### 3.2 Specification

Our estimation strategy is straightforward. First, we examine contemporaneous effects of teacher-student gender matches by estimating a series of specifications with the following general form:

$$y_{ijbgs} = \beta_0 + \beta_1 fs_i + \beta_2 ft_j + \beta_3 fs_i ft_j + X_i \delta'_1 + T_j \delta'_2 + \gamma_{bgs} + \varepsilon_{ijbgs}, \quad (1)$$

where  $y_{ijbgs}$  is a test score of student  $i$  taught by teacher  $j$  for subject  $b$  in ability group level  $g$ , if any, in 7th grade at school  $s$ .  $fs_i$  and  $ft_j$  are indicator

Table 3: Resampling Tests

	Math	English Language	Korean Language
<i>A. Test for Student Assignment</i>			
Kolmogorov-Smirnov test (No. failed / total tests)	0/74	0/74	0/74
$\chi^2$ goodness of fit test (No. failed / total tests)	1/74	1/74	4/74
<i>B. Test for Teacher Assignment</i>			
Female teacher	-0.017 (0.038)	-0.048 (0.061)	-0.022 (0.047)
Graduate school degree	-0.062 (0.035)	-0.020 (0.043)	-0.049 (0.039)
Full time teacher	-0.002 (0.048)	0.043 (0.056)	-0.062 (0.062)
Administrative teacher	0.030 (0.042)	-0.029 (0.067)	0.033 (0.045)
Observations	582	617	598

Each column represents a separate regression. The dependent variable is the empirical p-value from resampling. Independent variables are indicators for teacher gender, graduate degree, full time teacher, and administrative teacher. Each specification controls for school fixed effects. Standard errors in parentheses are clustered at school level. The Kolmogorov-Smirnov and  $\chi^2$  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.

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

variables equaling one if student  $i$  and subject teacher  $j$  are female, respectively. Our simplest specification includes only these variables. We add, in turn, an increasingly rich set of controls. These include  $X_i$ , a vector of students' predetermined characteristics including indicators for living with both parents, having at least one parent with bachelor's degree or higher, and having both parents being employed and  $T_j$ , a vector of teacher characteristics, including teacher's age and indicators for teacher graduating from teacher's school, teacher with master's degree, homeroom teacher, and teacher holding an administrative position at school. We include school by subject by ability group level fixed effects  $\gamma_{bgs}$  to compare students of the same ability in a subject within a school, to ensure that ordinary least squares (OLS) produces unbiased estimates. Standard errors are clustered at the school level to account for correlations among students within the same schools. If the random assignment suggested by the institutional features truly holds, these specifications should not differ substantially when the additional controls are included.

$\beta_1$  is the difference in average academic achievement between female and male students when taught by a male teacher.  $\beta_2$  represents the average difference in performance for male students between being taught by a female teacher and being taught by a male teacher.  $\beta_3$  indicates the change in the gender gap



between female and male students when switching from a male to a female teacher.

We modify Equation 1 slightly to examine the effects of teacher-student gender matches in 7th grade on standardized test scores over time:

$$y_{ijbgt_s} = \beta_0 + \beta_1 f s_i + \beta_2 f t_j + \beta_3 f s_i f t_j + \gamma_{bgs} + \varepsilon_{ijbgt_s}, \quad (2)$$

where  $y_{ijbgt_s}$  is test score in year  $t = 2, 3, 4,$  or  $5$  (namely, 8th through 11th grades), which is normalized within a subject and the year. We focus on the effects of 7th grade teachers over time because we are most confident that random assignment holds for that grade. Students are entering a randomly-assigned middle school and their academic performance cannot yet have been affected by teachers at the middle school when assigned to 7th-grade classrooms. This also provides the longest time span to examine the persistence of these effects.

Our primary interest is the coefficient  $\beta_3$  when  $t = 2, 3, 4,$  and  $5$ , representing gender gap effects one, two, three, and four years after exposure to the 7th grade teacher. When we examine outcomes that do not differ across subject, such as choice of academic track, we collapse this specification to a single observation per student.

## 4 Results

### 4.1 Contemporaneous Effects

We begin by examining the effect of teacher-student gender matches in 7th grade on students' standardized test scores in that year. Table 4 begins with the most parsimonious specification that includes no additional controls. The coefficient on the female student  $\times$  female teacher indicator in Column (1) indicates that the performance gap between female and male students increases by 13.3% of a standard deviation (s.e. = 6.8%) when the teacher is female rather than male. This effect consists of a statistically insignificant decrease in boys' performance of 0.06 standard deviations and an increase in girls' performance of 0.07. In other words, the gender gap effect is composed of an opposite-gender teacher effect for switching from a male teacher to a female one for male students ( $-\beta_2$ ) and a same-gender teacher effect for the switch to a female teacher for female students ( $\beta_2 + \beta_3$ ). This effect is substantial, considering the evidence that 10 additional days of schooling increases academic performance by 0.01 standard deviation (Carlsson et al., 2015). Adding controls across specifications does not meaningfully affect the estimate of this gender gap effect, providing additional evidence that random assignment holds in our data. Note, though, that the share of the gender gap effect attributable to changes in boys' performance does change. Boys no longer see a negative impact of switching to a female teacher by Column (3), but the estimates of  $\beta_2$  are noisy throughout.

Even the inclusion of student and teacher fixed effects in Columns (6) and (7), thus identifying from within-student and within-teacher variation in the

gender match, leaves the gender gap effect intact. Taken together, these results indicates that the teacher-student gender match is uncorrelated with observable and unobservable student and teacher characteristics. These findings replicate those using other South Korean data to examine same-year effects (Lim and Meer, forthcoming).<sup>4</sup>

Table 4: Contemporaneous Effects in 7th Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Student	0.069 (0.061)	0.128** (0.061)	0.087* (0.045)	0.095 (0.057)	0.117* (0.063)		
Female Teacher	-0.061 (0.078)	-0.044 (0.121)	0.063 (0.077)	0.054 (0.084)	0.039 (0.093)	0.002 (0.114)	
Female Student $\times$ Female Teacher	0.133* (0.068)	0.145** (0.070)	0.143*** (0.054)	0.141** (0.064)	0.132* (0.069)	0.175* (0.091)	0.180** (0.075)
Observations	10,194	10,194	10,045	8,470	7,792	9,217	10,045
$R^2$	0.008	0.161	0.370	0.391	0.381	0.827	0.823
Sch $\times$ Sbj FEs		Yes					
Sch $\times$ Sbj $\times$ Abg FEs			Yes	Yes	Yes	Yes	Yes
Student Controls				Yes	Yes		
Teacher Controls					Yes	Yes	
Student FEs						Yes	Yes
Teacher FEs							Yes

Each column shows results from a separate regression with normalized seventh grade test score as the dependent variable. Student controls include indicators for parents living together, both parents working, and at least one parent having college degree or higher. Teacher controls include indicators for having less than five years' experience, graduating from a teachers' school, and having a graduate degree. In Columns 6 and 7, female student and female teacher dummies are subsumed by student fixed effects and teacher fixed effects, respectively. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

## 4.2 Longer Term Effects

Table 5 presents the impacts of the 7th grade teacher-student gender match for later years. Panel A presents results including school-by-subject-by-ability-group fixed effects and Panel B also includes student and teacher fixed effects (corresponding to Columns (3) and (7), respectively, of Table 4). For brevity, we limit ourselves to the most parsimonious specification that we believe clearly satisfies the identifying assumptions (Panel A), but also present the most restrictive (Panel B).

Somewhat surprisingly, we find that the gender gap effects persist even four years after the initial teacher-student gender match. The effects vary slightly over time, but there are no significant differences between the contemporaneous

<sup>4</sup>That paper uses data on ninth graders throughout South Korea in 2004 and finds a contemporaneous interaction effect of female students with female teachers of about 0.10.

effect and the effects in the following years <sup>5</sup>. While education interventions are generally characterized by large fade-out effects (Jacob, Lefgren and Sims, 2010), previous studies on longer-run impacts of teacher gender suggest that the impacts of this particular phenomenon may be long-lasting (Carrell, Page and West, 2010; Lavy and Sand, 2015).

Table 5: Effects over Time

	Dep. Var. = Standardized Test Scores in			
	8th Grade (1)	9th Grade (2)	10th Grade (3)	11th Grade (4)
<i>A. Controls for Sch × Sbj × Grp FEs</i>				
Female Student	0.100 (0.061)	0.180*** (0.065)	0.135* (0.069)	0.090 (0.067)
Female Teacher in 7th Grade	-0.066 (0.064)	-0.112* (0.057)	-0.041 (0.061)	-0.108 (0.071)
Female Student × Female Teacher in 7th Grade	0.136** (0.061)	0.103 (0.067)	0.154** (0.067)	0.155** (0.062)
Observations	9,663	9,206	7,386	7,356
$R^2$	0.342	0.336	0.283	0.217
<i>B. Controls for Sch × Sbj × Grp FEs, Stu FEs, &amp; Tch FEs</i>				
Female Student × Female Teacher in 7th Grade	0.162** (0.069)	0.146 (0.093)	0.230*** (0.078)	0.217*** (0.079)
Observations	9,663	9,206	7,386	7,356
$R^2$	0.836	0.852	0.839	0.818

Each column shows results from a separate regression. Dependent variables in Columns 1 through 4 are normalized test scores within a subject and a year, in 8th through 11th grades, respectively. Panel A controls for 7th grade school by subject by 7th grade ability group level fixed effects and Panel B adds 7th grade student fixed effects and 7th grade teacher fixed effects. In Panel B, female student and female teacher dummies are subsumed by student fixed effects and teacher fixed effects, respectively. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

We next turn to advanced coursetaking in high school. In Table 6, we examine the impact of the 7th grade teacher-student gender match on advanced coursetaking in 11th grade. For this and subsequent results, we include school by subject by ability group effects. The results indicate that female students who had a female teacher in a given subject are about 8 percentage points more likely to take an advanced course in that same subject four years later.

STEM outcomes are of particular interest, since although the gender gap in math achievement in secondary education is small, women are substantially underrepresented in both STEM majors and careers. Column (1) of Table 7 presents the results for the effects of 7th grade subject teachers on student's

<sup>5</sup>In Panel A, the p-values are  $p = 0.92$ ,  $p = 0.52$ ,  $p = 0.91$ , and  $p = 0.89$  for the differences between the interaction coefficients in Columns (1)-(4) and that in Column (3) of Table 4, respectively. In Panel B, p-values are  $p = 0.83$ ,  $p = 0.72$ ,  $p = 0.64$ , and  $p = 0.74$  for the differences between the interaction coefficients in Columns (1)-(4) and Column (7) of Table 4

Table 6: Effects on Advanced Coursetaking

	Coefficient	S.E.
Female Student	-0.073**	(0.030)
Female Teacher in 7th Grade	-0.055*	(0.030)
Female Student $\times$ Female Teacher in 7th Grade	0.080**	(0.032)
Constant	0.608***	(0.027)
Observations	6,427	
$R^2$	0.303	

We regress an indicator for whether an 11th grader took at least one advanced course in a subject on student gender, 7th grade subject teacher gender and its interaction, controlling for 7th grade school by subject by 7th grade ability group level fixed effects. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

Table 7: Effects on STEM Outcomes

	(1) Math Track	(2) Advanced Math	(3) Plan STEM	(4) STEM School	(5) No Plan
FS	-0.177 (0.108)	-0.038 (0.106)	-0.124 (0.158)	-0.138 (0.112)	0.137* (0.075)
Math FT in 7th Grade	0.100 (0.069)	0.060 (0.053)	0.031 (0.088)	-0.038 (0.034)	0.076* (0.040)
Eng FT in 7th Grade	-0.019 (0.118)	0.196 (0.138)	-0.115 (0.168)	-0.021 (0.034)	0.128 (0.087)
Kor FT in 7th Grade	0.096 (0.080)	-0.014 (0.068)	0.146 (0.092)	0.014 (0.065)	0.054 (0.042)
FS $\times$ Math FT in 7th Grade	0.051 (0.060)	0.097* (0.055)	0.146* (0.080)	0.070* (0.041)	-0.065* (0.034)
FS $\times$ Eng FT in 7th Grade	-0.008 (0.068)	-0.203** (0.080)	-0.082 (0.121)	-0.001 (0.036)	-0.100* (0.053)
FS $\times$ Kor FT in 7th Grade	-0.068 (0.084)	0.056 (0.074)	-0.116 (0.096)	0.065 (0.093)	-0.096* (0.056)
Observations	1,635	1,456	1,043	1,652	1,640
$R^2$	0.229	0.248	0.280	0.259	0.267
Dep. Var. Mean	0.351	0.201	0.357	0.081	0.096

Each column represents a separate regression, controlling for school by math ability group by English ability group by Korean ability group in 7th grade fixed effects. Dependent variables are indicators for the choice of the math-science track in 11th grade, taking at least one advanced math course required for admission to a STEM postsecondary degree program, self-reporting the desire to seek a STEM postsecondary degree conditional on having decided a major, attending a STEM-oriented high school, and having no postsecondary plans. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

math-science track choice in 11th grade. These specifications include fixed effects for school by subject and 7th grade ability group, which should adequately control for any possible nonrandom assignment to ability group in 7th grade. There are no significant gender *gap* effects for any subjects; however, there is a statistically significant *same*-gender teacher effect for female students. Girls who had a 7th grade math teacher are 15.1 percentage points (*s.e.* = 7.5) more likely to chose the math-science track in high school when taught by female versus male math teacher (the combination of having a female math teacher and the additional effect of female math teachers on girls). Substantial gender gap effects emerge in advanced math coursetaking in 11th grade (Column 2), with having a female 7th grade math teacher reducing the gender gap between female and male students by 9.7 percentage points (*s.e.* = 5.5). The same gender teacher effect for female student is even bigger; female students are 15.7 percentage points (*s.e.* = 6.6) more likely to take at least one advanced math course when they were taught by female math teacher in 7th grade versus a male teacher. Column (3) shows the impacts on self-reported interest in majoring in a STEM field; once again, having a female 7th grade math teacher substantially reduces the male-female gap in these aspirations. Similarly, Column (4) shows that the male-female gap in attending a STEM-focused high school is substantially reduced when a female student has a female math teacher in 7th grade. Especially given the baseline means, listed in the bottom of line of the table, these effects are quite meaningful. Finally, Column (5) reports the impact on the likelihood that a student reports that they have no set plans for postsecondary education; female students are significantly less likely to report that they have no such plans if they had female teachers in any subject in 7th grade. This provides further evidence of an overall positive influence of female teachers on female students, which we turn to in the next section.

### 4.3 Evidence on Mechanisms

Altogether, these results show dramatic impacts of having female teachers on female students in the long run, with substantial increases in preparation for postsecondary STEM study. We examine several possible drivers of these results.

We begin by examining whether female students taught by a female teacher are more likely to be taught by a female teacher in later years. In that case, the persistent effects of 7th grade teachers could, in fact, be driven by cumulative effects of later teachers. As seen in Table 8, our results are not driven by cumulative exposure to female teachers set in motion by better performance in 7th grade.

We next examine whether teacher-student gender matches in 7th grade affects ability group assignment in later years. Table 9 shows results for the likelihood of being in a high ability group in 8th through 10th grades in a given subject. While this potential mechanism has a reasonable basis – better performance in 7th grade leads to higher ability grouping in later years, with better performance in those years – we find no evidence that this is driving our results.

Table 8: Effects on Future Teacher Gender

	Dep. Var. = Likelihood of Having Female Teacher in		
	8th Grade (1)	9th Grade (2)	10th Grade (3)
Female Student	0.026 (0.023)	-0.012 (0.018)	0.198*** (0.040)
Female Teacher in 7th Grade	0.034 (0.027)	-0.037 (0.022)	0.046 (0.038)
Female Student $\times$ Female Teacher in 7th Grade	-0.021 (0.024)	0.015 (0.019)	-0.041 (0.039)
Constant	0.806*** (0.023)	0.836*** (0.019)	0.429*** (0.034)
Observations	9,660	7,603	6,825
$R^2$	0.475	0.546	0.170

Each column represents a separate regression controlling for 7th grade school by subject by 7th grade ability group level fixed effects. Dependent variables in Columns 1 through 3 are indicators for having a female teacher in 8th through 10th grades. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

Table 9: Effect on Ability Grouping

	Dep. Var. = Likelihood of Being in High Ability Group in		
	8th Grade (1)	9th Grade (2)	10th Grade (3)
Female Student	-0.018 (0.028)	0.049 (0.048)	-0.050 (0.043)
Female Teacher in 7th Grade	-0.018 (0.040)	-0.086 (0.069)	-0.047 (0.056)
Female Student $\times$ Female Teacher in 7th Grade	0.016 (0.029)	-0.009 (0.051)	0.020 (0.048)
Constant	0.495*** (0.034)	0.539*** (0.057)	0.440*** (0.047)
Observations	3,633	1,711	3,494
$R^2$	0.450	0.464	0.202

Each column represents a separate regression, controlling for 7th grade school by subject by 7th grade ability group fixed effects. The response variables for Columns 1 through 3 are indicators for being in a high ability group level in 8th through 10th grades, respectively. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

Finally, we examine whether attrition from the sample is likely to bias our results. If attrition is systematically related to the teacher-student gender interaction – if low-performing female students with female teachers are more likely to exit the sample – our results would be the product of a selected sample. In Table 10, we test whether the proportion of female teachers in 7th grade impacts

the likelihood of attriting from the sample. No meaningful patterns emerge that indicate that selective attrition drives our results. Further, recreating our primary results in Tables 1 and 2 using only students who remained in the sample all five years yields similar results.

Table 10: Attrition

	Dep. Var. = Likelihood of Attrition in			
	8th Grade (1)	9th Grade (2)	10th Grade (3)	11th Grade (4)
FS	-0.009 (0.026)	0.052 (0.052)	0.035 (0.083)	-0.032 (0.049)
Pct. FTs in 7th Grade	0.038 (0.036)	-0.041 (0.045)	0.092 (0.120)	-0.077 (0.070)
FS × Pct. FTs in 7th Grade	0.014 (0.029)	-0.053 (0.063)	-0.108 (0.105)	0.034 (0.059)
Constant	0.007 (0.029)	0.075** (0.035)	0.138 (0.096)	0.103* (0.055)
Observations	2,292	2,204	2,113	1,721
$R^2$	0.114	0.186	0.293	0.196

Each column represents a separate regression, controlling for school by math ability group × English ability group × Korean ability group in 7th grade fixed effects. The dependent variables for Columns 1 through 4 are, respectively, indicators for a student attriting from the sample in 8th through 11th grades. The percent female teachers is defined the proportion of females in three subject of math, English, and Korean, conditional on no missing values in teacher gender. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

Having addressed these possibilities, we turn to student attitudes and choices. Indeed, our results appear to be driven by changes in student attitudes, which result in female students who had female teachers in 7th grade being more likely to select higher-quality high schools. Column (5) of Table 7 shows that having female teachers in 7th grade increases the likelihood that female students declare that they have postsecondary plans. In Table 11, we show the impact of the female student-female teacher match on an index of self-reported student engagement constructed using principal component analysis. While female students have a lower engagement score than male students, having a female teacher in a particular subject in 7th grade eliminates this gap, and the effects persist into high school.

Given the nature of the high school application process in Seoul, any increase in high school quality necessarily reflects student preferences to a great degree. In Table 13, we show the impact of the 7th grade teacher-student gender match on the quality of the student’s classroom peers, as estimated by previous-year standardized test scores, and on teacher value-added.<sup>6</sup> Given the random assignment in middle schools, it is reassuring that there are no effects in 8th and

<sup>6</sup>We calculate value added by estimating teacher fixed effects from a regression of leave-out mean scores in grade  $g$  on leave-out mean scores in grade  $g - 1$  and school-by-subject-by-ability-group fixed effects.

Table 11: Effects on Student Engagement Index

	Dep. Var. = Student Mechanism Index in				
	7th Grade (1)	8th Grade (2)	9th Grade (3)	10th Grade (4)	11th Grade (5)
FS	-0.107*** (0.035)	-0.083** (0.041)	-0.117** (0.058)	-0.108** (0.051)	-0.046 (0.073)
FT in 7th Grade	0.002 (0.052)	0.038 (0.071)	-0.044 (0.056)	-0.099* (0.053)	-0.063 (0.089)
FS $\times$ FT in 7th Grade	0.126*** (0.045)	0.103** (0.050)	0.144** (0.069)	0.146** (0.060)	0.068 (0.087)
Constant	0.004 (0.042)	-0.028 (0.057)	0.035 (0.046)	0.079* (0.043)	0.052 (0.072)
Observations	9,934	9,535	9,242	7,520	7,372
$R^2$	0.106	0.075	0.088	0.064	0.065

Each column represents a separate regression, controlling for school by subject by ability group in 7th grade fixed effects. Dependent variables are student behavior indices for 7th through 11th grades, constructed in the following way: first, we extract common variation, using principal component analysis, for each subject and each year, from the student’s response to 4 questions: whether he or she is focused in class, whether they participate in class, whether they do not skip homework assignments, and whether they review what they learned in class. The first component for each subject is regressed on student fixed effects; the residual is the student engagement index. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

9th grades. However, in 10th grade – high school – female students who had a female teacher in 7th grade have both peers and teachers who are significantly higher quality as measured by test scores.<sup>7</sup>

To complement the results on peer and teacher quality, we examine other measures of high school quality in Table 12. Column 1 shows the likelihood that the student attends a prestigious first-round high school, namely, an application-only high school, excluding non-elite vocational schools. The impact on the gender gap of having more female teachers is statistically insignificant. In Column 2, there is a substantial but statistically insignificant effect on the likelihood of attending a high school in a different administrative district, which would reflect a willingness to incur greater costs to attend a preferred school. This effect is not entirely driven by students who attend first-round schools. Examining only those who attend second-round (lottery) high schools, the female student interaction with percent of 7th grade female teachers is 0.074 (s.e. = 0.078); once again, insignificant but fairly sizable. This suggests even those female students with more female teachers who do not attend an application school are requesting higher-quality schools in the second-round lottery.

Column 3 shows the impacts on the percent of 11th grade students in a high school who achieved “Above Basic” performance in a particular subject

<sup>7</sup>Due to data limitations discussed in Section 2.2, we are unable to estimate these results for 11th grade.



Table 12: Effects on High School Quality

	(1) Prestigious 1st Round HS	(2) Different District	(3) Above Basic	(4) Below Basic	(5) Pct. Violent Students
FS	-0.108 (0.080)	-0.112 (0.090)	0.005 (0.008)	-0.002 (0.003)	0.002 (0.003)
Pct. FT in 7th Grade	0.001 (0.107)	0.060 (0.121)			0.002 (0.004)
FS × Pct. FT in 7th Grade	0.031 (0.088)	0.098 (0.118)			-0.007* (0.003)
FT in 7th Grade			-0.015* (0.009)	0.007*** (0.002)	
FS × FT in 7th Grade			0.035*** (0.011)	-0.010*** (0.004)	
Constant	0.180** (0.086)	0.163* (0.092)	0.799*** (0.007)	0.048*** (0.002)	0.009*** (0.003)
Observations	1,721	1,719	6,576	6,580	1,705
$R^2$	0.269	0.200	0.265	0.296	0.375
Dep.Var.Mean	0.142	0.197	0.801	0.049	0.009
Sch × Math Grp × Eng Grp × Kor Grp FEs	Yes	Yes			Yes
Sch × Subj × Grp FEs			Yes	Yes	

Notes: Each column represents a separate regression. Columns 3 and 4 are pooled data for all three subjects and control for school by subject by ability group in 7th grade fixed effects. In Columns 1, 2, and 5, we collapse the outcomes and control for for school by math ability group by English ability group by Korean ability group in 7th grade fixed effects. The dependent variables in Columns 1 and 2 are indicators for attending a prestigious first round high school (i.e. one of the first round high schools excluding non-elite vocational high schools) and a high school located in the different administrative district from the student's residential district. The dependent variable in Column 3 is the percent of 11th grade students with Above Basic Performance in National Assessment of Educational Achievement (NAEA) test; in Column 4, the percent of 11th grade students with Below Basic Performance in NAEA test. The NAEA exam scores are for the cohort one year ahead of the students in our sample. The dependent variable in Column 5 is the percent of students who are reported to School Violence Committee while our students in sample are in 10th grade. Students are reported for actions committed against other students inside or outside of school premises resulting in a physical or mental injury, or damage to property through a battery, assault, confinement, threat, kidnapping, abduction, defamation, insult, extortion, coercion, forced errand, sexual violence, bullying, or cyber-bullying, or spreading obscene or violent information via a communications network. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

Table 13: Effects on Peer and Teacher Quality in Later Years

Dep. Var. =	Peer Quality in			Teacher Quality in		
	8th Grade (1)	9th Grade (2)	10th Grade (3)	8th Grade (4)	9th Grade (5)	10th Grade (6)
FS	-0.010 (0.029)	0.013 (0.014)	0.071 (0.047)	-0.011 (0.017)	0.009 (0.009)	0.041 (0.046)
FT in 7th Grade	-0.028 (0.022)	0.015 (0.012)	-0.045 (0.058)	0.002 (0.020)	-0.011 (0.016)	-0.029 (0.038)
FS $\times$ FT in 7th Grade	-0.002 (0.029)	-0.011 (0.014)	0.152*** (0.053)	0.014 (0.015)	-0.001 (0.010)	0.085** (0.041)
Observations	9,615	9,028	7,095	9,578	8,961	6,766
$R^2$	0.656	0.793	0.279	0.749	0.719	0.597

Each column represents a separate regression, controlling for school by subject by ability group fixed effects. The dependent variables in Columns 1 to 3 are peer quality measures and those in Columns 4 through 6 are teacher quality measures. Peer quality is defined as the mean standardized test score in previous year, excluding a student's own score, for each subject teacher. Teacher quality is the teacher fixed effect in 8th through 10th grades, obtained as follows: for each teacher  $j$  in  $g$ th grade (where  $g = 8, 9$  and  $10$ ), take all the students whom teacher  $j$  teaches. Then, for each  $g$ th grade student  $i$  under teacher  $j$ , calculate the leave-out means (excluding student  $i$ ) for  $(g - 1)$ th and  $g$ th grade test scores. Next, regress the leave-out mean for  $g$ th grade on that for  $(g - 1)$ th grade and  $g$ th grade teacher fixed effects, controlling for 7th grade school  $\times$  subject  $\times$  7th grade ability group fixed effects. Standard errors in parentheses are clustered at school level.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

on the National Assessment of Educational Achievement (NAEA) test; this indicates achievement above the median grade-level expectation. Importantly, these scores are for students one cohort *above* those in our sample. Column 4 shows results for students performing at the “Below Basic” level, below 20 percent of grade-level expectations. Finally, Column 5 shows the effects on the percent of students who are reported to the School Violence Committee, to which infractions against students both on and off school premises must be reported; these include violent acts but also bullying and cyber-bullying. Female students with a greater proportion of female teachers in 7th grade attend safer schools, though we note, at least based on official reports, schools in South Korea are extremely safe.

Taken together, these results indicate that the student-teacher gender interaction in middle school leads to greater engagement for female students with female teachers. Since the impact on attending application-only schools is fairly small, the primary channel for these results must be the choice set submitted by students for the second-round lottery. That is, exposure to female teachers in 7th grade leads female students to express a preference for higher-quality schools, even at a cost of greater travel time.

## 5 Conclusions

Our study shows that female students' pairing with female teachers at a formative age – middle school – increases standardized test scores well into high school, the likelihood of advanced course-taking and making plans for postsecondary attendance, and, specifically for those with female math teachers, both studying STEM and aspirations to do so in college. We show that persistent student engagement and increased ambition in choosing high schools of higher quality explain why there is little fade-out of these impacts. It is important to note that there are, at times, partially offsetting negative effects on academic outcomes for boys from a shift to female teachers. Overall, these findings shed light on the importance of teacher-student gender matches in future policies to close the gender gap, especially in STEM fields.

## References

- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik.** 2015. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics*, 33(1): 63–89.
- Berenbaum, Sheri A., Carol Lynn Martin, and Diane N. Ruble.** 2008. "Gender Development." *Child and Adolescent Development: An Advanced Course*, 647–695.
- Byun, Sooyong, and Kyungkeun Kim.** 2010. "Educational inequality in South Korea: The widening socioeconomic gap in student achievement." *Research in Sociology of Education*, 17(155-182).
- Carlsson, Magnus, Gordon B Dahl, Björn Öckert, and Dan-Olof Rooth.** 2015. "The effect of schooling on cognitive skills." *Review of Economics and Statistics*, 97(3): 533–547.
- Carrell, Scott E., and James E. West.** 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409–432.
- Carrell, Scott E., Marianne E. Page, and James E. West.** 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics*, 125(3): 1101–1144.
- Cho, Insook.** 2012. "The Effect of Teacher Student Gender Matching: Evidence from OECD Countries." *Economics of Education Review*, 31(3): 54–67.
- Dee, Thomas S.** 2007. "Teachers and the gender gaps in student achievement." *Journal of Human Resources*, 42(3): 528–554.
- Good, Phillip I.** 2006. *Resampling Methods: A Practical Guide to Data Analysis*. 3rd ed. Boston: Birkhäuser.
- Holmlund, Helena, and Krister Sund.** 2008. "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics*, 15(1): 37–53.
- Jacob, Brian A, Lars Lefgren, and David P Sims.** 2010. "The persistence of teacher-induced learning." *Journal of Human resources*, 45(4): 915–943.
- Kang, Changhui.** 2007. "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea." *Journal of Urban Economics*, 61(3): 458–495.
- Korea Legislation Research Institute.** 2011. "Enforcement Decree of the Elementary and Secondary Education Act, Article 68 (Methods for Entering Middle Schools)." *Presidential Decree No.22712*.

- Lavy, Victor, and Edith Sand.** 2015. "On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers Stereotypical Biases." *NBER Working Paper No. 20909*.
- Lehmann, Erich L., and Joseph P. Romano.** 2005. *Testing Statistical Hypotheses (Springer Texts in Statistics)*. . 3rd ed., Springer.
- Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland.** 2015. "Expectations of brilliance underlie gender distributions across academic disciplines." *Science*, 347(6219): 262–265.
- Lim, Jaegeum, and Jonathan Meer.** forthcoming. "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea." *Journal of Human Resources*.
- Muralidharan, Karthik, and Ketki Sheth.** 2016. "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers." *Journal of Human Resources*, 51(2): 269–97.
- Seoul Metropolitan Office of Education.** 2012. "Announcement on High School Admissions Plan for 2013 School Year." *SMOE Announcement No.2012-51*.
- 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.