The Effect of Teacher Gender on Student Achievement in Primary School

Heather Antecol, *Claremont McKenna College and IZA*

Ozkan Eren, *Louisiana State University*

Serkan Ozbeklik, *Claremont McKenna College*

Using data from a randomized experiment, we find that having a female teacher lowers the math test scores of female primary school students in disadvantaged neighborhoods. Moreover, we do not find any effect of having a female teacher on male students' test scores (math or reading) or female students' reading test scores, which seems to rule out explanations pertaining to the unobserved quality differences between male and female teachers. Finally, this negative effect seems to persist only for female students who were assigned to a female teacher with a limited math background.

I. Introduction

A number of recent studies in economics have attempted to document the effect of having a female teacher on different academic outcomes, especially performance in math and the choice of a math and science major,

We wish to thank Kelly Bedard, Scott Carrell, Thomas Dee, Philip Oreopoulos, Marianne Page, and all seminar participants at the 2013 Allied Social Science Associations annual meetings, the 2012 All CA Labor Conference, and the University of California, Santa Barbara, for their helpful comments and suggestions. Contact the corresponding author, Heather Antecol, at hantecol@cmc.edu. Information concerning access to the data used in this article is available as supplementary material online.

[*Journal of Labor Economics,* 2015, vol. 33, no. 1]
 0734-306X/2015/3301-0003$10.00
Submitted November 27, 2012; Accepted August 9, 2013; Electronically published October 31, 2014

of female students, either in middle school/high school (e.g., Ehrenberg, Goldhaber, and Brewer 1995; Nixon and Robinson 1999; Dee 2005, 2007; Winters et al. 2013) or postsecondary education (e.g., Canes and Rosen 1995; Rothstein 1995; Neumark and Gardecki 1998; Bettinger and Long 2005; Hoffman and Oreopoulos 2009; Carrell, Page, and West 2010).[1] These studies either find that having a female teacher has a positive effect on female student achievement outcomes (e.g., Rothstein 1995; Nixon and Robinson 1999; Bettinger and Long 2005; Dee 2007; Hoffman and Oreopoulos 2009; Carrell et al. 2010; Winters et al. 2013) or no effect on them (e.g., Canes and Rosen 1995; Ehrenberg et al. 1995; Neumark and Gardecki 1998).[2]

The impact of teacher gender on the outcomes of students in primary school has received less attention to date. There has, however, been significant media coverage drawing attention to a growing concern that primary school female teachers have an adverse impact on female students but not male students (see, e.g., Kaplan 2010; Mack 2010; and Molina 2010). This stylized fact is based on a recent study in the educational psychology literature, which finds that having a female primary school teacher leads to lower math test scores among female students but not male students (see Beilock et al. 2010).[3] In contrast, a recent study in the economics literature finds no evidence of a relationship between teacher gender and test score outcomes (math and reading) of students, irrespective of gender, in primary school (see Winters et al. 2013).[4]

[1] Although the focus of this paper is gender interactions within classrooms, there are several studies in the economics literature investigating the effects of similarities in gender and ethnicity on the academic achievement of students. See, e.g., Ehrenberg et al. (1995), Dee (2004), and Fairlie, Hoffman, and Oreopoulos (2011).

[2] We are aware of only one study, by Dee (2007), that finds that being assigned to a female teacher is associated with lower math test scores for female students in eighth grade. However, after conducting several robustness checks, and given that he also finds a similar effect for male students, Dee concludes that this is largely due to the nonrandom assignment of female teachers to classrooms with low-performing students in math. For the remainder of the analysis, Dee does not focus on math achievement; he solely focuses on achievement in English, history, and science. Dee finds that a female teacher has a large positive effect on history outcomes for female students. He also finds smaller positive effects in English and science, but these effects are not statistically significant at conventional levels.

[3] Beilock et al. (2010) also show that the more anxious female primary school teachers are in math classes and the more likely female students are to endorse the stereotypes “boys are good at math, and girls are good at reading,” the lower the math achievement of female students relative to male students or female students without such a belief. We investigate this potential mechanism, henceforth referred to as the “math anxiety hypothesis,” further in Sec. III.C.

[4] Winters et al. (2013) also examine the effect of teacher gender on male and female student test score outcomes in grades 6–10. They find that female teachers have a positive impact on the test scores (math and reading) of students, irrespec-

We argue that it is important to further our understanding of the impact of teacher gender on primary student outcomes, particularly in light of the different results in the two strands of literature, for the following reasons. The teacher-student gender dynamics in primary school might be different than they are for higher levels of education. In particular, the gender differences in children's self-perceptions about ability and their awareness of commonly held beliefs about gender stereotypes start emerging between the ages of 7 and 12 (Eccles et al. 1993; Steele 2003). Moreover, primary school experiences may shape the academic course of students, leading to long-term gender differences in fields of study and occupational choices, which in turn can lead to long-term differences in earnings capacity. While casual empiricism suggests that the gender gap in math test scores occurs at higher levels of education, recent evidence suggests that the math test scores of boys and girls begin to diverge in primary school across developed countries (see Bedard and Cho [2010] and references therein).[5]

Furthermore, most of the existing literature could not account for nonrandom assignment of teachers, which may cause their estimates to be biased, including both Bielock et al. (2010) and Winters et al. (2013).[6] Specifically, if high-achieving and better-motivated students are less likely to be assigned to female teachers, then the effect of having a female teacher (relative to a male teacher) on achievement will be understated. In addition, the analysis in Bielock et al. (2010) is based on a very small sample (17 teachers, 65 female students, and 52 male students) from one urban school district (one school) in the Midwest, and these researchers are unable to examine the relative effectiveness of male and female teachers on the achievement outcomes of male and female students due to an insufficient sample of male teachers.[7] Finally, Winters et al. (2013) appears to be based on a very small number of male teachers in grades 3–5 in Florida only,

tive of gender, in these higher grades. Given that they find an effect for both male and female students, their results may be due to the nonrandom assignment of female teachers to classrooms with high-performing students in math or average-quality differences between male and female teachers in their data.

[5] There is, however, disagreement in the US literature regarding the exact age/grade at which the gender gap in math test scores occurs. In particular, Coley (2001) and Bedard and Cho (2010) document a gap as early as grade 4, Freeman (2004) documents a gap as early as grade 3 but no such gap in grade 1, while Dee (2007) finds no evidence of a gap for 9-year-olds.

[6] To our knowledge, in the existing economic literature, only Carrell et al. (2010) use experimental data from the US Air Force Academy, which is a very selective postsecondary institution, to analyze the effect of teacher gender on grades in math and science courses, as well as the probability of taking a higher-level math course and the probability of graduating with a degree in science, technology, engineering, and/or mathematics (STEM).

[7] Female teachers generally constitute about 90% of all teachers in primary schools (Bursal and Pagnozas 2006; Gresham 2007).

and it does not account for classroom fixed effects, nor does it formally test the differential effect of a female teacher on male and female student test score outcomes.

In this paper, we attempt to fill these gaps in the literature using data from a well-executed randomized experiment that was conducted to evaluate the effectiveness of the Teach for America (TFA) Program.[8] Specifically, we look at the effects of having a female teacher (relative to a male teacher) on the math test scores of female students in primary schools in disadvantaged neighborhoods.[9] We also analyze the effects of having a female teacher on the reading test scores of female students and the math and reading test scores for male students. Finally, we investigate the potential mechanisms that might help explain the influence of teacher gender on primary student outcomes.

Our unique data set affords us several advantages over the existing literature. First, the experimental nature of our data, which comes from 17 schools in six different states, allows us to avoid the issue of nonrandom assignment of teachers. Second, the data also allow us to control for either block fixed effects (classrooms in the same school and the same grade) or classroom fixed effects, which allows us to identify gender differences between teachers and their students within each block/classroom. Third, our data afford us a large sample of primary schools, students, teachers, and states. Moreover, our data come from a very disadvantaged part of the student population. This allows us to take a closer look at the teacher-student interactions in a setting where the problems with the education system in the United States are most evident and arguably are especially important from a policy perspective. Specifically, one of the goals of the Race to the Top (RTTT) Program is to improve schools that have the lowest achievement levels. Not surprisingly, the lowest-achievement schools tend to be located in disadvantaged neighborhoods (see Antecol, Eren, and Ozbeklik [2013] and references therein for a more detailed discussion). Finally, our data afford us a large sample of male teachers, allowing us to examine the relative effectiveness of male and female teachers on the achievement outcomes (math and reading) of male and female students.

[8] Teach for America (TFA) is a nonprofit organization that recruits outstanding recent college graduates and mid-career professionals to teach in schools in highly disadvantaged neighborhoods throughout the United States for at least a 2-year period. Since its inception in 1990, approximately 24,000 TFA Corps Members have taught more than 3 million students in 38 urban and rural areas. Between 2000 and 2009 the number of applications for TFA skyrocketed to 35,000 from 4,068 and the number of new corps members recruited each year grew from 868 to 4,100. See http://www.teachforamerica.org for more detailed information about the organization.

[9] This is an artifact of the randomized experiment given that members of the TFA program mostly teach in disadvantaged neighborhoods.

We find that female students who were assigned to a female teacher, as opposed to a male teacher, suffered from lower math test scores at the end of the academic year. We do not find any effect of having a female teacher on male students' test scores (math or reading) or female students' reading test scores. These robustness checks seem to rule out the explanations pertaining to the unobserved quality differences between male and female teachers. Finally, this negative effect seems to persist only for female students who were assigned to a female teacher with a limited math background.

Although our results are based on a small number of teachers with a strong math background, they extend the existing literature in a number of ways—we focus on disadvantaged neighborhoods, employ randomized data, include male teachers in the estimation sample, include block fixed effects or classroom fixed effects, and utilize reading test scores as a robustness check—and they provide suggestive evidence regarding the potential mechanisms for our results (this is discussed in detail in Sec. III.C).

II. Data

We use data from the Mathematica Policy Research, Incorporated (MPR), National Evaluation of Teach for America (NETFA) Public Use File. NETFA is a randomized study of primary school students in six regions in the United States between 2001 and 2003. The pilot study was conducted in Baltimore in the 2001–2 academic year, followed by full-scale evaluations in Chicago, Los Angeles, Houston, New Orleans, and the Mississippi Delta in the next academic year (2002–3). Each region had one school district participating in the experiment, with the exception of the Mississippi Delta, which had two school districts that participated in the experiment. Within each school district, schools were selected to reflect where TFA placed teachers at the time of the study, and only schools that had both TFA teachers and "control" teachers in the same grade were considered eligible for the study. These schools are generally disadvantaged and face substantial teacher shortages; thus, they are not a representative sample of average schools in the United States. For example, across the schools in the TFA study, the average rate of student eligibility for a free or a reduced-priced lunch was about 97% as opposed to 41% nationwide. The final sample consisted of 17 schools, 100 teachers, and more than 1,900 students in grades 1–5.[10]

These data are ideal for our purposes because all students were randomly assigned to two types of classes, TFA and control group classrooms, before the start of the academic year. Therefore the randomization is done at the block level, such that each block represents classrooms

[10] See Decker, Mayer, and Glazerman (2004) for further details about this experiment and its data.

in the same grade level in any given school. Furthermore, throughout the year roster checks were performed to enforce the original assignment. These processes ensured not only that those students in TFA and control group classrooms are comparable but also that the gender of the students and the gender of their teacher in each classroom are not correlated (this is discussed in further detail in Sec. III.A). After the random assignment and before the start of academic year, the students were given math and reading tests based on the grade they had completed in the previous academic year (which we call *pretreatment* outcome variables); then at the end of the academic year in which the study was conducted, the students retook math and reading tests based on the grade they had just completed (which we call *posttreatment* outcome variables).[11]

Although sample attrition is relatively small, we lose around 14% (16%) of our initial math (reading) test score sample because of missing test scores, missing teacher characteristics, and students moving out of the school districts. After dropping these observations, our estimation sample consists of 1,664 (1,624) students for the math (reading) test score sample from classes taught by 95 teachers.[12] To ensure that the student composition was unaffected by the sample attrition, we show that student characteristics, as well as teacher characteristics (if available), are similar (i.e., statistically indistinguishable) across the randomization sample and our estimation sample (see table 1).[13]

[11] Reading test scores are based on the combined scores of vocabulary and word analysis for all grades except grade 1. Due to data limitations, grade 1 reading test scores are based on vocabulary only. Results for reading are similar if we exclude grade 1 from the analysis and are available upon request.

[12] Our estimation samples (1,664 and 1,624 observations for the math and reading test score sample, respectively) are slightly different than the full sample Mathematica researchers use (1,715 observations) because several teachers had missing information for control variables related to their characteristics necessary for our estimation. Despite this, we are able to replicate the results of Mathematica researchers with respect to the effect of TFA teachers on both math and reading test scores (with only slight differences in the coefficient estimates). These results are available upon request.

[13] We exclude two classrooms (31 students) for whom we do not have information on teacher gender in both the randomization sample and the estimation sample. Our analysis includes all classrooms irrespective of whether the classroom had more than one teacher during the academic year (i.e., seven out of the 98 [95] classrooms in the randomization [estimation] sample had multiple teachers). In all but one of these seven classrooms only one teacher responded to the survey. In the classroom where more than one teacher responded to the survey, we based the teacher information on the first teacher who taught in the classroom; for all other classrooms with multiple teachers, we based the teacher information on the teacher who responded to the survey (i.e., one classroom is based on the first teacher who taught in the classroom, three classrooms are based on the second teacher who taught in the classroom, one classroom is based on the third teacher who taught in

We also investigate whether the sample attrition differs between the classrooms taught by male teachers and the classrooms taught by female teachers. Specifically, we regress the sample attrition dummy on a teacher gender dummy, student characteristics, block fixed effects, and a TFA indicator variable. We find that sample attrition did not differ by teacher gender. The coefficient on the teacher gender dummy is always small and has a large standard error irrespective of sample, that is, the coefficient (standard error) for the whole randomization sample is 0.035 (0.033) and 0.015 (0.031) for the sample without missing pretreatment math test scores.[14] Given that sample attrition does not appear to be an issue, for the remainder of the discussion of descriptive statistics, we focus on our estimation sample.

A. Student Characteristics

The data include detailed information on student characteristics: type of class, gender, race and ethnicity, math test scores (pre- and posttreatment), reading test scores (pre- and posttreatment), and class size (see col. 1 of table 1 for all variable definitions). There are two types of classes to which a student is randomly assigned: a class taught by a TFA teacher or a class taught by a control teacher. On average, 44% of the students (based on the math test score sample) were assigned to classrooms taught by a TFA teacher (see col. 3 of table 1), and this assignment is not significantly different for male and female students.[15]

We consider three racial/ethnic groups for students: non-Hispanic black, Hispanic, and non-Hispanic white, henceforth referred to as black, Hispanic, and white, respectively.[16] Given that the schools in our sample are disadvantaged schools, the student body is predominantly nonwhite. Specifically, 66.6% (27.3%) of the student body is black (Hispanic), while 6.1% is white. Furthermore, the average number of students per class is 25.1, and the mean pretreatment (baseline) normal curve equivalent (NCE) math test score is 30.2, whereas the mean pretreatment (baseline) NCE reading test score is 28.6. Students in our sample are about one standard deviation below the national average (i.e., the NCE scale has a mean of 50 and a standard deviation of 21 nationally).

the classroom, and one classroom is based on the fourth teacher who taught in the classroom). We find similar results if we exclude the seven classrooms (154 students) with multiple teachers during the academic year. Results are available upon request.

[14] We also find that attrition does not vary between students in the TFA and control group classrooms. For the sake of brevity, we do not report these results in the paper, but they are available upon request.

[15] Similar results are found based on the reading test score sample.

[16] The white sample includes a small number of Asians and Pacific Islanders.

Table 1
Descriptive Statistics by Sample

	Definition (1)	Randomization Sample (2)	Estimation Sample (3)	Difference (2) − (3)
Student characteristics:				
Female	1 if student female, 0 otherwise	.489	.493	−.005
		(.011)	(.012)	(.017)
White	1 if student non-Hispanic white or Asian/Pacific Islander, 0 otherwise	.069	.061	.008
		(.006)	(.006)	(.008)
Black	1 if student non-Hispanic black, 0 otherwise	.671	.666	.005
		(.011)	(.012)	(.016)
Hispanic	1 if student Hispanic, 0 otherwise	.260	.273	−.013
		(.010)	(.011)	(.015)
Class size	Current number of students in the classroom	25.104	25.090	.014
		(.129)	(.137)	(.188)
Pretreatment math*	Average national curve equivalent (NCE) pretreatment math test score	29.646	30.214	−.568
		(.429)	(.457)	(.627)
Pretreatment reading†*	Average NCE pretreatment reading test score	28.767	28.622	.145
		(.445)	(.471)	(.649)
Teacher characteristics:				
Female	1 if teacher female, 0 otherwise	.764	.759	.005
		(.010)	(.010)	(.014)
TFA	1 if Teach for America teacher, 0 otherwise	.436	.435	.000
		(.011)	(.012)	(.017)

Certified	1 if teacher has traditional teacher certification, 0 otherwise		.551 (.012)
Teacher experience	Years of teaching experience		6.392 (.203)
White	1 if teacher non-Hispanic white or Asian/Pacific Islander, 0 otherwise		.400 (.012)
Black	1 if teacher non-Hispanic black, 0 otherwise		.480 (.012)
Hispanic	1 if teacher Hispanic, 0 otherwise		.101 (.007)
Strong math background	1 if teacher majored/minored in math or math-related (computer science, system analysis, engineering systems, premed, economics, and accounting) subject in college/post-college, 0 otherwise		.114 (.008)
No. of observations		1,938	1,664

SOURCE.—Author's calculations using the Mathematica Policy Research, Inc. (MPR) National Evaluation of Teach for America (NEFTA) Public Use File, 2001–2003.

NOTE.—Table presents means with standard errors reported (in parentheses), except where otherwise indicated. Neither sample includes two classrooms (31 students) for whom we do not have information on teacher gender.

* Pretreatment reading test scores are based on slightly smaller samples for the estimation sample due to nonreporting; 1,624 students with pretreatment reading test scores.

† Pretreatment test scores are based on slightly smaller samples for the randomized sample due to nonreporting; 1,880 and 1,881 students with pretreatment math test scores and reading test scores, respectively.

B. Teacher Characteristics

The data also include detailed information on teacher characteristics. Specifically, we have information on the teachers' gender, race/ethnicity, certification status, teaching experience, major in college, and major in their graduate degree if they have one. About one-quarter of the students in our sample are taught by male teachers (see col. 3 of table 1).[17] As previously noted, this allows us to examine the relative effectiveness of male and female teachers and to overcome a shortcoming of earlier research on teacher effectiveness in primary schools due to insufficient samples of male teachers.[18] Roughly 48% of the students have black teachers, 10% have Hispanic teachers, and 40% have white teachers.[19] For the remaining 3%, we do not have information on the race/ethnicity of the teacher.

Teachers were asked to describe what type of certification, license, and credential they hold. We consider two broad certification types: regular certification (including standard state certificate) and nonregular certification (including emergency, temporary, initial, and other types of certifications). About 55% (45%) of the students had teachers with regular (nonregular) certification.[20] On average, students had a teacher with slightly more than 6 years of experience.

Finally, in terms of college majors the teachers are very diverse. Since we are concerned with the potential mechanisms that might help explain the influence of teacher gender on primary student math outcomes, we construct an indicator for whether a teacher has a strong math background (see Sec. III.C for a detailed discussion). As such, we focus our attention on one particular set of college (undergraduate or graduate) majors/minors: math or math-related (i.e., computer science, system analysis, engineering systems, premed, economics, and accounting) majors/minors. Teachers who

[17] In our sample, about 33% of TFA teachers are male (i.e., 13 out 41 TFA teachers are male) as opposed to slightly less than 15% of control teachers (i.e., eight out of 54 control teachers are male). Since the characteristics of TFA teachers are quite different than those of regular teachers in our sample (see app. table A1), we control for whether a teacher is a TFA teacher or a control teacher in all our regressions to prevent any confounding effect this difference might generate.

[18] Higher levels of education do not suffer from the same shortage of male teachers, and thus this is generally not a shortcoming of the economics literature on teacher effectiveness.

[19] The white category again includes a small number of Asian/Pacific Islander teachers. Female teachers are less likely to be white than their male counterparts; this difference is largely driven by TFA teachers rather than control teachers (see app. table A1).

[20] Female teachers are more likely to have regular certification relative to male teachers. Interestingly, this difference is driven by control teachers rather than TFA teachers (see app. table A1).

have a strong math background taught roughly 11.4% of the students in our sample.

III. Threats to Identification, Estimation Strategy, and Results

A. Threats to Identification and Validity of Randomized Data

Any study analyzing the effect of teachers on student achievement has to deal with two important potential identification problems that might bias the conventional ordinary least squares (OLS) estimates. First, schools assign students to teachers nonrandomly, even within subjects and grade levels (e.g., Clotfelter, Ladd, and Vikdor 2006; Kane et al. 2011). In the current context, if high-achieving and better-motivated students are more likely to be assigned to male teachers, then the effect of having a male teacher on achievement will be overstated. The common practice in the literature to deal with nonrandom sorting is to control for students' prior achievement (e.g., Hanushek and Rivkin 2010). In a recent paper, however, Rothstein (2010) shows that the estimates of teachers' contribution on student achievement may still be biased even after conditioning on prior achievement. Second, even if students and teachers are randomly assigned to each other, there may be unobserved teacher traits that are correlated with student outcomes that again may bias the conventional OLS estimates, for example, unobserved gender-specific differences across teachers' quality.

Fortunately, the randomization of our data at the classroom level allows us to avoid the identification problems associated with nonrandom assignment of students and teachers. This is because, as we previously explained, students were randomly assigned to each teacher at the beginning of the academic year, and regular class roster checks throughout the year ensured that class compositions did not change. Therefore, nonrandom sorting of students to teachers of their parents' choice was avoided.[21] The second problem, on the other hand, is harder to deal with directly. As we

[21] Of course it would be interesting to show how important the nonrandom selection is for our results if we had comparable nonexperimental data on primary schools. While we do not have direct evidence, we can point to the analysis by Dee (2007), who argues that the nonrandom assignment of female teachers to classrooms with low-performing students in math can explain why he finds that being assigned to a female teacher is associated with lower math test scores for female students in the eighth grade. Arguably, controlling for nonrandom assignment is important; however, as we outlined earlier, our analysis extends the existing literature on primary schools in a number of other ways beyond employing randomized data (i.e., we focus on disadvantaged neighborhoods, include male teachers in the estimation sample, and utilize reading test scores as a robustness check); thus, we cannot disentangle the relative importance of nonrandom selection from these other extensions.

explain in detail below, however, we use different specifications and robustness checks to show that potential unobserved teacher traits do not appear to have a significant influence on our results.

In order to illustrate that the randomization generated comparable students by teacher gender, we regress the teacher gender dummy on student characteristics (controlling for a TFA indicator variable and block fixed effects) using all students in the experiment with information on the gender of their teacher (see col. 1 of table 2). As pretreatment math and reading test scores are unavailable for all students, these measures are excluded from this specification. We do, however, estimate two additional specifications, one based on the sample of students with pretreatment math test scores, which includes a control for pretreatment math test scores (see col. 2 of table 2), and an analogous regression for the sample of students with pretreatment reading test scores (see col. 3 of table 2). This reduces the sample size by 58 (57) observations for the sample of students with pretreatment math (reading) test scores.

If the randomization worked, we would expect to see small and statistically insignificant coefficients for each of these characteristics irrespective of sample. As expected, none of coefficients are statistically significant at conventional levels, and in general the p-values are very large.[22] Taken together, these results show that randomization ensured that teacher gender and students characteristics are not correlated.

There are also some drawbacks to using experimental data. One potential problem is if students and/or parents know that they are a part of an experiment, they may behave differently than they would do otherwise. In the current context, this should not be a problem since the main focus of the experiment was to assess the effect of TFA teachers (not female teachers) relative to control teachers (not male teachers). Hence, the effects associated with the participants' knowledge that they belong to a particular experimental assignment should not be relevant. Another potential drawback is the external validity of the experiment. As we men-

[22] While we do find some evidence of a difference in the overall student composition of classrooms taught by male and female teachers, if we exclude race/ethnicity, this difference disappears (see table 2). We also run separate regressions of each individual student characteristic on teacher gender, a TFA indicator variable, and block fixed effects. As expected, the coefficient on teacher gender is always extremely small irrespective of student characteristic, and it is generally statistically insignificant at conventional levels (the one exception is that the coefficient on teacher gender in the black indicator regression is marginally significant, p-value of .092). We argue that it is not uncommon to find some differences by chance when doing this type of testing for randomization. This argument is further reinforced by the fact that we find no evidence of a difference between the pretreatment math and reading test scores of black (and Hispanic) students in classrooms taught by female teachers relative to classrooms taught by male teachers.

Table 2
Test for Randomization: Dependent Variable Is Female Teacher Dummy

	All Students (1)	All Students with Pretreatment Math Test Scores (2)	All Students with Pretreatment Reading Test Scores (3)
Female student	−.010	−.010	−.005
	(.008)	(.009)	(.008)
Class size	.005	.003	−.004
	(.018)	(.019)	(.023)
Black student	−.136	−.133	−.131
	(.118)	(.125)	(.126)
Hispanic student	.103	.090	.092
	(.134)	(.134)	(.133)
Pretreatment math		.001	
		(.001)	
Pretreatment reading			−.001
			(.001)
Joint test without race/ethnicity	.900	.610	.950
(F-statistic [p-value])	[.415]	[.612]	[.425]
Joint test with race/ethnicity	2.670	2.24	2.440
(F-statistic [p-value])	[.048]	[.072]	[.053]
Block fixed effects	Yes	Yes	Yes
No. of observations	1,938	1,880	1,881
R^2	.421	.425	.439

SOURCE.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.
NOTE.—Table presents coefficients and standard errors (in parentheses), except where otherwise indicated. We control for a TFA indicator variable in all regressions given TFA status is the treatment variable in the original experiment, as well as an indicator variable for Hispanic student missing. The null hypothesis in the joint significance test is that all the variables displayed in the table are not jointly significant. Standard errors are clustered at the block level. See table 1 for variable definitions.

tioned in the preceding section, the data come from very disadvantaged parts of the US population, such that almost 97% of the students are eligible for the free lunch program and that there are very few white students (slightly more than 6%). Therefore, we do not claim that our results can necessarily be generalized to the entire US student population, but, as previously noted, they allow us to focus on a sample where the problems with the education system in the United States are well documented.

B. Estimation Strategy and Results

In order to formally determine the effect of female teachers on student achievement in primary school, we estimate a regression of the following form:

$$\begin{aligned} \mathrm{TS}_{icb} &= \beta_0 + \beta_1 \mathrm{FEM}_{icb} + \beta_2 \mathrm{FEMTEACH}_{icb} \\ &+ \beta_3(\mathrm{FEM}_{icb} \times \mathrm{FEMTEACH}_{icb}) + \mathrm{SC}'_{icb}\delta + \mathrm{TC}'_{icb}\gamma + \alpha \mathrm{TFA}_{icb} + \eta_b + \varepsilon_{icb}, \end{aligned} \tag{1}$$

where TS is subject-specific post-treatment student test scores for student i in classroom c and block b (i.e., classrooms in the same school and the same grade), and FEM is equal to one if the student is female and zero otherwise, FEMTEACH is equal to one if the student's teacher is female and zero otherwise, SC is a vector of student characteristics (i.e., race/ethnicity, pretreatment NCE math or NCE reading test scores, and class size), TC is a vector of teacher characteristics (i.e., race/ethnicity, years of teaching experience and its square, and type of certification), and TFA is equal to one if the student's classroom was taught by a TFA teacher and zero otherwise. Finally, η represents block fixed effects, and ε is an error term with the standard properties.[23] This specification implies that essentially our identification is coming from gender differences between teachers and their students within each block.[24] Moreover, we use the fact that teacher gender may differentially affect female and male students. This allows us to estimate an effect of teacher gender on the relative test score outcomes of female and male students even when there may be unobserved differences between the test score outcomes of all students in classrooms taught by female teachers versus classrooms taught by male teachers (e.g., due to unmeasured differences in teacher quality).

Table 3 presents our results for posttreatment NCE math and NCE reading test scores; henceforth, we refer to these as math and reading test scores. Following Dee (2007), we present all our regression results for specifications with and without teacher characteristics after controlling for student characteristics, a TFA indicator variable, and block fixed effects. As the results are similar for the models with and without teacher characteristics, for the sake of brevity, we focus on the models with the full set of controls in the remainder of the discussion (cols. 2 and 6 of table 3).[25]

We find that female students relative to male students who were taught by a female teacher ($\beta_1 + \beta_3$) fare significantly worse (i.e., score 1.9 points lower, which is 10% of a sample standard deviation) on their math test scores. Moreover, female students relative to male students who were taught by a male teacher (β_1) have higher math test scores; however, the effect is insignificant at conventional levels (p-value of .38). The difference-in-differences coefficient (β_3), which gives the relative difference in math test scores between female students with female versus male teachers and male students with female versus male teachers, is large (−3.35, which is a little less than 20% of a sample standard deviation) and statistically significant at the 10% level. Finally, we find no significant effects of teacher gender

[23] The standard errors are clustered at the block level.

[24] We also estimate a specification that uses differences in teacher and student gender within each classroom to identify our parameters of interest as an additional robustness check (see detailed discussion below).

[25] We obtain similar results if we exclude NCE equivalent scores of zero from the analysis for both math and reading. These results are available upon request.

Table 3
Determinants of Math and Reading Test Scores: Block Fixed Effects

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female student (β_1)	1.500	1.460	2.166	2.131	1.648	1.727	1.984	2.055
	(1.659)	(1.655)	(1.605)	(1.604)	(1.651)	(1.661)	(1.766)	(1.780)
Female teacher ($\beta 2$)	−.638	−.408	.186	.260	−.345	−.360	.211	.098
	(1.737)	(2.062)	(1.952)	(2.190)	(1.441)	(1.353)	(1.620)	(1.546)
Female student × Female teacher (β_3)	-3.393^{+}	-3.347^{+}	−4.379*	−4.351*	−1.173	−1.303	−1.681	−1.794
	(1.875)	(1.869)	(1.896)	(1.893)	(1.651)	(1.658)	(1.815)	(1.832)
Strong math background ($\beta 4$)			2.230	1.955			2.190	.010
			(4.732)	(5.512)			(1.381)	(1.943)
Female teacher × Strong math background ($\beta 5$)			−6.109	−5.509			-4.383^{+}	−3.482
			(5.273)	(6.076)			(2.491)	(2.798)
Female student × Strong math background ($\beta 6$)			−5.998*	−5.958*			−4.044	−4.050
			(2.662)	(2.661)			(2.835)	(2.858)
Female student × Female teacher × Strong math background (β_7)			9.144**	9.106**			5.660	5.671
			(3.165)	(3.199)			(4.529)	(4.568)
$\beta_1 + \beta_3$	−1.893*	−1.887*			.475	.425		
	(.733)	(.728)			(.851)	(.847)		
$\beta_3 + \beta_7$			4.765^{+}	4.755			3.979	3.877
			(2.786)	(2.843)			(3.665)	(3.686)
Outcome variable mean	32.136 [19.709]				29.638 [18.484]			
Teacher characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	1,664	1,664	1,664	1,664	1,624	1,624	1,624	1,624
R^2	.550	.551	.552	.552	.562	.565	.563	.566

SOURCE.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.

NOTE.—Table presents coefficients and standard errors (in parentheses), except where otherwise indicated. We control for a TFA indicator variable in all specifications. Student characteristics include race/ethnicity, pretreatment NCE math or NCE reading scores, and class size. Teacher characteristics include race/ethnicity, years of teaching experience and its square, and type of certification. Standard errors are clustered at the block level. Standard deviations are in square brackets. See table 1 for variable definitions.

$^{+}$ $p < .10$.
* $p < .05$.
** $p < .01$.

on reading achievement of male and female students.[26] The fact that female teachers appear to only lower the math achievement of female students and not male students, and not the reading achievement of either male or female students, suggests that the estimated effect of female teachers on math achievement is not biased by unobservable confounders.

Taken together, our results suggest that female primary school teachers (relative to male teachers) adversely influence the math outcomes of female students but not male students. While one could alternatively interpret our results as male primary school teachers (relative to female teachers) positively influence the math outcomes of female students but not male students, we argue the former interpretation is more likely given that we continue to find that the adverse effect of female teachers on the math outcomes of female students persists even when we exclude male teachers from the analysis (see app. tables A2 and A3).

In order to ensure that our results are not driven by unobservable differences between male and female teachers in the same grade and the same school, we reestimate equation (1) controlling for classroom fixed effects instead of block fixed effects. Specifically, we estimate an equation of the following form:

$$\mathrm{TS}_{ic} = \delta_0 + \delta_1 \mathrm{FEM}_{ic} + \delta_2(\mathrm{FEM}_{ic} \times \mathrm{FEMTEACH}_{ic}) + \mathrm{SC}'_{ic}\lambda + \mu_c + e_{ic}, \tag{2}$$

where μ represents classroom fixed effects and all other variables are as previously defined.[27] Our results from the classroom fixed effects specification (see cols. 1 and 3 of table 4) are very similar to our results from the block fixed effects specification (see cols. 2 and 6 of table 3) in terms of both the magnitudes of the coefficients and the significance levels. This provides further evidence that the patterns we are finding are not an artifact of differences across our sample of teachers.

Overall our results seem to confirm the results found in Bielock et al. (2010)—female primary school teachers (relative to male teachers) adversely influence the math outcomes of female students but not male students—and they seem to be at odds with Winters et al. (2013), who find no

[26] In order to ensure that our results are not driven by blocks that do not include both male and female teachers, we rerun our analysis for blocks that include both male and female teachers, thereby excluding blocks with only female (18 blocks) and male teachers (1 block). We are left with 18 blocks out of a total of 37 blocks, and the sample is roughly halved. Our results from this exercise are very similar and are available upon request. Ideally we would have reestimated models that included controls for a teacher's math background (see Sec. III.C); however, the small number of teachers with a math background (i.e., four teachers) in these remaining 18 blocks did not afford us this option.

[27] The standard errors are clustered at the classroom level.

Table 4
Determinants of Math and Reading Test Scores: Classroom Fixed Effects

	Math		Reading	
	(1)	(2)	(3)	(4)
Female student (δ_1)	1.312	2.165	1.533	1.897
	(1.655)	(1.748)	(1.797)	(1.944)
Female student × Female teacher (δ_2)	-3.174^{+}	-4.390^{*}	−1.061	−1.579
	(1.826)	(1.934)	(1.950)	(2.091)
Female student × Strong math background (δ_3)		-7.478^{**}		−4.336
		(2.682)		(2.779)
Female student × Female teacher × Strong math background (δ_4)		10.637^{**}		5.647
		(3.116)		(4.008)
$\delta_1 + \delta_2$	-1.863^{*}		.472	
	(.789)		(.774)	
$\delta_2 + \delta_4$		6.247^{*}		4.068
		(2.428)		(3.404)
Outcome variable mean	32.136	[19.709]	[29.638]	[18.484]
Student characteristics	Yes	Yes	Yes	Yes
Classroom fixed effects	Yes	Yes	Yes	Yes
No. of observations	1,664	1,664	1,624	1,624
R^2	.580	.582	.580	.580

SOURCE.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.
NOTE.—Table presents coefficients and standard errors (in parentheses), except where otherwise indicated. Student characteristics include race/ethnicity and pretreatment NCE math or NCE reading scores. Standard errors are clustered at the classroom level. Standard deviations are in square brackets. See table 1 for variable definitions.
$^{+}$ $p < .10$.
* $p < .05$.
** $p < .01$.

such effect. Our results likely differ from Winters et al. (2013) because we focus on a very different sample of students from disadvantaged neighborhoods in grades 1–5 across six states (i.e., Maryland, Illinois, California, Texas, Louisiana, and Mississippi), we use random assignment, and we control for either block or classroom fixed effects. Winters et al. (2013), on the other hand, focus on students in primary school (grades 3–5) in Florida only using a first-difference estimator based on a 5-year administrative-level panel data set. As such, the magnitudes of the effect of teacher gender on student achievement outcomes in their analysis may potentially be biased as they are based on nonrandomized data and do not control for classroom fixed effects. Further evidence to suggest this is the case can be found if one looks at the results they present on the effect of teacher gender on the test score outcomes of students in middle/high school (grades 6–10). Specifically, they find that test scores (reading and math) are positively influenced by female teachers for students, irrespective of gender, in these higher grades. Given that they find an effect for both male and female students, their results may be due to the nonrandom assignment of female teachers to classrooms with high-performing students in math or average-quality differences between male and female teachers in their data.

C. Potential Mechanisms for the Math Achievement Outcomes of Female Students

In an attempt to explore the potential mechanisms for female student math outcomes, we control for the strength of a teacher's math background. We posit that the negative effect on math achievement due to being assigned to a female teacher may disappear for the female students taught by female teachers with a strong math background.

Specifically, math teaching styles between female teachers with and without a strong math background may vary, and these differences may affect the math performance of male and female students differently. For example, it is possible to obtain this result if female teachers without a strong math background are worse at teaching math due to improper training (e.g., chose to teach in a mechanical way) that is fine for male students but harmful for female students at this age. We argue, however, that this is not likely to be the case since, to the best of our knowledge, there is no evidence either on these potential differences in teaching styles or on whether these differences in teaching styles differentially influence male and female students.[28] Alternatively it could be that female teachers, with or without a strong math background, teach math as well as male teachers but that female students respond better to a male teacher at this age. This seems less plausible, however, given that we find that the adverse effect of a female teacher on female student math test scores persists even when we exclude male teachers from the analysis (see app. tables A2 and A3). Finally, it could be that a female teacher with a strong math background is less likely to suffer from math anxiety. This may not only be a result of better math knowledge and ability but may also indicate that the female teacher is less likely to hold stereotypical gender beliefs about math.

While we cannot distinguish between the potential mechanisms given our estimation strategy, we argue that "the math anxiety hypothesis" is more plausible in light of the existing evidence in the educational psychology literature. In particular, Beilock et al. (2010) find that higher math anxiety in female primary school teachers hurts the math performances of the female students but not of the male students in the first and second grades.[29] Moreover, they show that this negative effect works through fe-

[28] While we do have some information on teaching styles and teacher beliefs in our data set, this information is not student-gender specific. We find that there are some important differences in teacher beliefs between male and female teachers, but these differences in beliefs do not result in differences in teaching styles between male and female teachers.

[29] Beilock et al. (2010) measure teacher math anxiety through a questionnaire that was given to teachers at the beginning and the end of the academic year.

male students' beliefs about who is good at math; the more anxious female teachers are in math classes and the more likely female students are to endorse the stereotype "boys are good at math, and girls are good at reading," the lower the math achievement of female students relative to both male students and female students without such a belief.[30]

In order to control for a teacher's math background with our data, we run the following regression specification:

$$\begin{aligned} TS_{icb} = {} & \beta_0 + \beta_1 FEM_{icb} + \beta_2 FEMTEACH_{icb} \\ & + \beta_3 (FEM_{icb} \times FEMTEACH_{icb}) + \beta_4 MATHREL_{icb} \\ & + \beta_5 (FEMTEACH_{icb} \times MATHREL_{icb}) \\ & + \beta_6 (FEM_{icb} \times MATHREL_{icb}) \\ & + \beta_7 (FEM_{icb} \times FEMTEACH_{icb} \times MATHREL_{icb}) \\ & + SC'_{icb}\delta + TC'_{icb}\gamma + \alpha TFA_{icb} + \eta_b + \varepsilon_{icb}, \end{aligned} \tag{3}$$

where MATHREL is equal to one if the classroom was taught by a teacher with a strong math background (i.e., a college/postcollege major/minor in math or a math-related subject—computer, system analysis, engineering systems, premed, economics, and accounting) and zero otherwise, and all other variables are as previously defined.[31] We must note that one

Specifically, math anxiety was assessed using the short Mathematics Anxiety Rating Scale, where teachers responded to 25 questions about how anxious different situations would make them feel (e.g., "reading a cash register receipt after you buy something," "studying for a math test," etc.). Responses were recorded on a Likert scale from 1 (low anxiety) to 5 (high anxiety). Beilock et al. (2010) construct an aggregate teacher math anxiety measure based on the average of the teacher responses from all 25 questions. Math anxiety among primary school teachers, of which females constitute about 90%, is a commonplace phenomenon (Bursal and Pagnozas 2006; Gresham 2007). Earlier studies also find a negative effect of teacher math anxiety on student performance in math classes (Tobias and Weissbrod 1980; Bush 1989); however, they do not discuss different effects for male and female students.

[30] Students in the study were told two stories, one about a student who was good at math and another about a student who was good at reading at the beginning and end of the year. They were then asked to draw these students. To determine the extent to which students endorsed the traditional stereotype "boys are good at math, and girls are good at reading," Beilock et al. (2010) focused on the gender of the students in the drawings and formed a measure. The higher the score, the more children ascribed to stereotypical gender roles in school. Note that the math anxiety argument is in the tradition of the stereotype threat model introduced in Steele (1997), who argued that a person can experience anxiety or concern in a situation where she or he has the potential to confirm a negative stereotype about the social group the person belongs to.

[31] The standard errors are clustered at the block level.

should be cautious in interpreting the results based on these regressions, since we only have 11 teachers with a strong math background in our sample (eight female teachers with 143 students and three male teachers with 46 students).

The results are presented in columns 4 and 8 of table 3.[32] For math test scores, the relative difference in the test scores between female students assigned to female versus male teachers and male students with female versus male teachers with a strong math background ($\beta_3 + \beta_7$) is a very large positive number (4.755), although insignificant at conventional levels. Similarly, the relative difference in the test scores between female students assigned to female versus male teachers and male students with female versus male teachers without a strong math background (β_3) is very large and negative (-4.351) and significant at the 5% level. Finally, the difference-in-difference-in-differences estimator (β_7) is very large and positive (9.106) and significant at the 1% level.[33] When we run the same regression for reading test scores as a robustness check, none of the coefficients of interest are statistically significant at conventional levels.[34]

While our estimates at first glance may seem high relative to the existing economics literature on higher levels of education, this may in part be due to the differences in the sample considered in our analysis, as well as the randomized nature of our data.[35] Specifically, in addition to focusing on students in primary schools, more than 90% of the students in our sample

[32] The results using classroom fixed effects are similar. See cols. 2 and 4 of table 4.

[33] We also restricted the analysis to both female teachers only (irrespective of TFA status) and female control teachers only and generally find similar patterns (although in some cases we lose precision). This suggests that the small number of male teachers with a math or math-related major do not appear to be biasing our results. See app. tables A2 and A3.

[34] We cannot rule out the possibility that the reason we do not find any effect for reading test scores pertains to having less power due to small sample sizes, particularly small samples of male teachers. While the magnitudes on the coefficients of interest in the reading test score regressions are somewhat large, they are substantially smaller than the coefficients in the math test score regressions. Thus, even if power is an issue, our results present strong evidence in favor of the detrimental effect of having a female teacher without a strong math background on math test scores of female students in primary school. Moreover, when we exclude male teachers, our results continue to support these findings. Finally, it is also possible that reading test scores are much less sensitive to teacher quality than math test scores.

[35] Dee (2007) finds that the negative effect of female teachers on the math test scores of eighth-grade students is roughly 6%–8% of a standard deviation, which is roughly half the size of the negative effect we find for female students who were taught by a female teacher (relative to a male teacher) without a strong math background. However, if we restrict our analysis to female teachers only, we find that this negative effect is 9%–10% of a standard deviation.

are from very disadvantaged minority families. One might expect the results to be different among primary school students given gender differences in children's self-perceptions about ability and that their awareness of commonly held beliefs about gender stereotypes starts emerging between the ages of 7 and 12 (Eccles et al. 1993; Steele 2003). Moreover, gender stereotypes and their effects on student outcomes may be more pronounced in disadvantaged families. As such, not having a female role model with a strong math background, especially in primary school, is potentially more detrimental for female students from disadvantaged families. Furthermore, the magnitudes of the effect of teacher gender on student achievement outcomes presented in the existing literature for higher levels of education may potentially be biased as they are generally based on nonrandomized data.

In addition, the effect of teacher gender on student achievement presented in the existing economics literature for higher levels of education (particularly postsecondary education) may indeed to be more closely aligned with the positive (although insignificant) effect we find for female students taught by female teachers with a strong math background as these studies tend to focus on select institutions of higher education with highly selected female teachers/professors who have extensive training in math and/or highly motivated female students who excelled in math at early ages may have self-selected into math-related fields of study. Thus, it is unclear whether our upper-bound estimates are indeed inconsistent with the existing economics literature.

Regardless of the magnitudes of our coefficient estimates, taken together our findings provide suggestive evidence that female students taught by female teachers without a strong math background have lower math achievement outcomes, and this does not appear to be driven by the small number of male teachers with a strong math background in our sample but by the nature of our sample and data.

IV. Conclusion

The impact of teacher gender on the outcomes of students in primary school has received little attention to date. In a recent study in the educational psychology literature, Beilock et al. (2010) finds that having a female teacher has a negative effect on female student math achievement in primary school.[36] In contrast, in a recent study in the economics literature, Winters et al. (2013) finds no evidence of a relationship between teacher gender and test score outcomes (math and reading) of students, irrespective of gender, in primary school.

[36] Earlier studies also find a negative effect of math anxiety on teaching performance in math classes (i.e., Tobias and Weissbrod 1980; Bush 1989); however, they do not discuss different effects for male and female students.

This paper further analyzes the effect of teacher gender on primary school student outcomes using evidence from a well-executed randomized experiment. Our unique data allow us to avoid the issue of nonrandom assignment of teachers and control for either block fixed effects (classrooms in the same school and the same grade) or classroom fixed effects. As well, it affords us a large sample of primary schools, students, teachers, and states. Furthermore, our data come from a very disadvantaged part of the student population, which allows us to take a closer look at the teacher-student interactions in a setting where the influence of gender stereotypes may be particularly pronounced and the problems with the education system in the United States are most evident and arguably more important from a policy perspective. Finally, our data provide us with a large sample of male teachers, allowing us to examine the relative effectiveness of male and female teachers on the achievement outcomes of male and female students.

Our results show that having a female teacher has a negative impact on the math test scores of female students in primary school in disadvantaged neighborhoods. Moreover, the negative impact of female teachers on the math achievement outcomes of female students does not appear to be an artifact of female and male teachers having differential unobserved characteristics, because we do not find a similar negative effect on the reading test scores of female students, nor do we find a negative effect on the test scores (math or reading) of male students. Finally, we find that the negative effect of having a female teacher on math test scores of female students seems to disappear for students taught by female teachers with a strong math background. Thus, only female students taught by female teachers with limited math backgrounds appear to be adversely affected.

Although this latter finding is based on a small number of teachers with a strong math background, taken together our results provide suggestive evidence in support of the math anxiety hypothesis found in the educational psychology literature (see Bielock et al. 2010), that is, that math anxiety among primary school female teachers in conjunction with female student endorsement of gender stereotypes may be leading to poorer math achievement among female students but not male students. Future research, with available randomized data including direct measures of math anxiety among female teachers and students' beliefs of gender stereotypes, as opposed to our proxy measure of the math background of the teacher, is needed to provide more definitive evidence of the math anxiety hypothesis. Despite this, the current analysis is a good first step at trying to understand why female teachers adversely influence the math test scores of female students but not male students in primary schools in disadvantaged neighborhoods.

Our findings also shed light on some policy prescriptions for primary school education. Research in educational psychology suggests that a mathematics methods course could lessen mathematics anxiety in prospective primary school teachers and improve the quality of classroom instruction (e.g., Hembree 1990; Emenaker 1996; Vinson 2001). Therefore, training programs focusing on method courses in math might be useful to counteract this negative impact, particularly for female teachers who teach in schools located in disadvantaged neighborhoods. Moreover, since this negative effect seems to potentially work through female students' beliefs about commonly held gender stereotypes, education policy geared at dispelling these beliefs (perhaps as early as preschool) would be a useful complementary policy that might have many other positive longer-term effects for female students as well as society in general.

We want to end by pointing out that our results apply to a very specific subset of the population, students in very disadvantaged neighborhoods. As such, it is unclear if our results can be generalized to the general population of primary school students. Having said this, however, this may be the subset of students we should be focusing on given that this is the setting where the problems with the education system in the United States are most evident and arguably more important from a policy perspective. Future research should address this important issue in further detail.

Appendix

Table A1
Teacher Characteristics by Treatment Status and Gender

	Overall			Control			TFA		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Race/ethnicity (%):									
Black	48.4	28.6	54.1*	70.4	62.5	71.7	19.5	7.7	25.0
White	38.9	57.1	33.8+	13.0	0	15.2	73.2	92.3	64.3+
Hispanic	10.5	14.3	9.5	14.8	37.5	10.9+	4.9	0	7.1
Race missing	3.2	0	4.1	3.7	0	4.3	2.4	0	3.6
Certification (%):									
Regular	55.8	28.6	63.5**	64.8	12.5	73.9**	43.9	38.5	46.4
Nonregular	44.2	71.4	36.5**	35.2	87.5	26.1**	56.1	61.5	53.6
Average years of teaching experience	6.6	4.1	7.3	10.0	7.8	10.4	2.0	1.8	2.0
% of teachers with a strong math background	11.6	14.3	10.8	11.1	25.0	8.7	12.2	7.7	14.3
No. of teachers	95	21	74	54	8	46	41	13	28

SOURCE.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.
NOTE.—See table 1 for variable definitions.
+ Significantly different than males; $p < .10$.
* Significantly different than males; $p < .05$.
** Significantly different than males; $p < .01$.

Table A2
Determinants of Math and Reading Test Scores for Female Teachers Only: Block Fixed Effects

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female student (λ_1)	−1.747*	−1.796*	−2.065*	−2.149*	.530	.496	.358	.318
	(.771)	(.775)	(.834)	(.848)	(.859)	(.857)	(.846)	(.854)
Strong math background ($\lambda 2$)			−4.531	−3.731			−1.911	−3.025
			(3.045)	(3.093)			(2.577)	(2.754)
Female student × Strong math background (λ_3)			3.114*	3.225*			1.567	1.598
			(1.452)	(1.506)			(2.875)	(2.895)
$\lambda_1 + \lambda_3$			1.048	1.076			1.925	1.917
			(1.310)	(1.331)			(2.838)	(2.842)
Outcome variable mean	32.397 [19.495]				29.348 [18.681]			
Teacher characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	1,263	1,263	1,263	1,263	1,236	1,236	1,236	1,236
R^2	.542	.548	.544	.549	.600	.602	.600	.602

SOURCE.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.

NOTE.—Table presents coefficients and standard errors (in parentheses), except where otherwise indicated. We control for a TFA indicator variable in all specifications. Student characteristics include race/ethnicity, pretreatment NCE math or NCE reading scores, and class size. Teacher characteristics include race/ethnicity, years of teaching experience and its square, and type of certification. Standard errors are clustered at the block level. Standard deviations are in square brackets. See table 1 for variable definitions.

* $p < .05$.

Table A3
Determinants of Math and Reading Test Scores for Female Control Teachers Only: Block Fixed Effects

	Math		Reading	
	(1)	(2)	(3)	(4)
Female student (λ_1)	−2.435*	−2.760*	.295	.112
	(1.060)	(1.079)	(1.000)	(.892)
Strong math background ($\lambda 2$)		−3.697		−1.767
		(3.564)		(3.986)
Female student × Strong math background (λ_3)		3.836		2.097
		(2.320)		(4.635)
$\lambda_1 + \lambda_3$		1.076		2.209
		(2.184)		(4.655)
Outcome variable mean	31.318 [18.798]		29.617 [18.040]	
Teacher characteristics	Yes	Yes	Yes	Yes
Student characteristics	Yes	Yes	Yes	Yes
Block fixed effects	Yes	Yes	Yes	Yes
No. of observations	789	789	775	775
R^2	.510	.511	.600	.600

Source.—Author's calculations using the MPR NEFTA Public Use File, 2001–2003.
Note.—Table presents coefficients and standard errors (in parentheses), except where otherwise indicated. Student characteristics include race/ethnicity, pretreatment NCE math or NCE reading scores, and class size. Teacher characteristics include race/ethnicity, years of teaching experience and its square, and type of certification. Standard errors are clustered at the block level. Standard deviations are in square brackets. See table 1 for variable definitions.
* $p < .05$.

References

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2013. The effect of Teach for America on the distribution of student achievement in primary school: Evidence from a randomized experiment. *Economics of Education Review* 37 (December): 113–25.

Bedard, Kelly, and Insook Cho. 2010. Early gender test score gaps across OECD countries. *Economics of Education Review* 29, no. 3:348–63.

Beilock, Sian L., Elizabeth A. Gunderson, Gerardo Ramirez, and Susan C. Levine. 2010. Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, USA* 107, no. 5: 1060–63.

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

Bursal, Murat, and Lynda Paznokas. 2006. Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science. *School Science and Mathematics* 106, no. 4:173–79.

Bush, William S. 1989. Mathematics anxiety in upper elementary school teachers. *School Science and Mathematics* 89, no. 6:499–509.

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

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

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41, no. 4:778–820.

Coley, Richard J. 2001. *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Princeton, NJ: Educational Testing Service.

Decker, Peter T., Daniel P. Mayer, and Steven Glazerman. 2004. The effects of Teach for America on students: Findings from a national evaluation. MPR Report 8792–8750, Mathematica Policy Research, Inc., Princeton, NJ.

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

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

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

Eccles, Jacquelynne, Allan Wigfield, Rena D. Harold, and Phyllis Blumenfeld. 1993. Age and gender differences in children's self- and task perceptions during elementary school. *Child Development* 64, no. 3: 830–47.

Ehrenberg, Ronald G., Dan D. Goldhaber, and Dominic J. Brewer. 1995. Do teachers' race, gender and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review* 48, no. 3:547–61.

Emenaker, Charles. 1996. A problem-solving based mathematics course and elementary teachers' beliefs. *School Science and Mathematics* 96, no. 2:75–84.

Fairlie, Robert, Florian Hoffman, and Philip Oreopoulos. 2011. A community college instructor like me: Race and ethnicity interactions in the classroom. NBER Working Paper no. 17381, National Bureau of Economic Research, Cambridge, MA.

Freeman, Catherine, E. 2004. *Trends in educational equity of girls and women: 2004 (NCES 2005–016)*. US Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.

Gresham, Gina. 2007. A study of mathematics anxiety in pre-service teachers. *Early Childhood Education Journal* 35, no. 2:181–88.

Hanushek, Eric A., and Steven G. Rivkin. 2010. Using value-added measures of teacher quality. *American Economic Review* 100, no. 2:267–71.

Hembree, Ray. 1990. The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education* 21, no. 1:33–46.

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

Kane, Thomas J., Eric S. Taylor, John H. Tyler, and Amy L. Wooten. 2011. Identifying effective classroom practices using student achievement data. *Journal of Human Resources* 46, no. 3:587–613.

Kaplan, Karen. 2010. Female teacher may pass on math anxiety to girls, study finds. *Los Angeles Times*, January 26.

Mack, Kristen. 2010. Study: Female teachers' math anxiety affects girl students. *Chicago Tribune*, January 25.

Molina, Brett. 2010. Girls may learn math anxiety from female teachers. *USATODAY*, January 25.

Neumark, David, and Rosella Gardecki. 1998. Women helping women? Role model and mentoring effects on female Ph.D. students in economics. *Journal of Human Resources* 33, no. 1:220–46.

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

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

Rothstein, Jesse. 2010. Teacher quality in educational production: Tracking, decay and student achievement. *Quarterly Journal of Economics* 25, no. 1:175–214.

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

Steele, Jennifer. 2003. Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology* 33, no. 12:2587–606.

Tobias, Sheila, and Carol Weissbrod. 1980. Anxiety and mathematics: An update. *Harvard Educational Review* 50, no. 1:63–70.

Vinson, Beth. 2001. A comparison of pre-service teachers' mathematics anxiety before and after a methods course emphasizing manipulatives. *Early Childhood Education Journal* 29, no. 2:89–94.

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