THE EFFECT OF TEACHER GENDER ON STUDENT ACHIEVEMENT IN PRIMARY SCHOOL

Heather Antecol*
Claremont McKenna College and IZA

Ozkan Eren
University of Nevada, Las Vegas

Serkan Ozbeklik
Claremont McKenna College

November 26, 2012

Abstract

Using data from a randomized experiment, we look at the effects of having a female teacher on the math test scores of primary school students in disadvantaged neighborhoods. We find that female students who were assigned to a female teacher suffered from *lower* math test scores at the end of the academic year. 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 not only seems to disappear but it becomes (marginally) positive for female students who were assigned to a female teacher *with* a strong math background. Thus, only female students taught by female teachers with *limited* math backgrounds appear to be adversely affected.

* Corresponding Author. All future correspondence regarding this manuscript may be addressed to Heather Antecol, Boswell Professor of Economics, The Robert Day School of Economics and Finance, Claremont McKenna College, 500 E. Ninth St., Claremont, CA 91711. Tel: (909) 607-8067. Fax: (909) 621-8249.
E-mail: hantecol@cmc.edu.

1. 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, of female students, either in middle school/high school (e.g., Ehrenberg et al. [1995], Nixon and Robinson [1999], Dee [2005], [2007]) or post-secondary education (e.g., Canes and Rosen [1995], Rothstein [1995], Neumark and Gardecki [1998], Bettinger and Long [2005], Hoffman and Oreopoulos [2009], Carrell et al. [2010]).[1] These studies either find 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]) 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, to the best of our knowledge, has received no attention to date in the economics literature. There has however been significant media coverage drawing attention to a growing concern that primary school female teachers adversely impact female students but not male students (see for example, 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

---

[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 for example, Ehrenberg et al. (1995), Dee (2007), and Fairlie et al. (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 8th grade. However, after conducting several robustness checks and given he also finds a similar effect for male students, he concludes this is largely due to the non-random assignment of female teachers to classrooms with low performing students in math. For the remainder of the analysis Dee (2007) 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.

leads to *lower* math test scores among female students but not male students (see Beilock et al. 2010).[3]

We argue that it is important to further our understanding of the impact of teacher gender on primary student outcomes 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.

Furthermore, most of the existing literature could not account for non-random assignment of teachers which may cause their estimates to be biased.[4] 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. Finally, 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 (1 school) in the Midwest and they are unable to examine the relative effectiveness of male and female teachers on the

---

[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 Section 3.3.

[4] To our knowledge, in the existing economic literature, only Carrell et al. (2010) use experimental data from the U.S. Air Force Academy, which is a very selective post-secondary 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).

achievement outcomes of male and female students due to an insufficient sample of male teachers.[5]

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.[6] 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 school in disadvantaged neighborhoods.[7] 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 6 different states, allows us to avoid the issue of non-random assignment of teachers. Second, our data affords us a large sample of primary schools, students, teachers, and states. Moreover, our data comes 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 more important from a policy perspective. Finally, our data affords 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.

---

[5] Female teachers generally constitute about 90 percent of all teachers in primary schools (Bursal and Pagnozas [2006], Gresham [2007]).
[6] TFA is a non-profit organization that recruits outstanding recent college graduates and mid-career professionals to teach in schools in highly disadvantaged neighborhoods for at least two-years throughout the United States. This program is explained in more detail in the next section.
[7] This is an artifact of the randomized experiment given 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. [8] These robustness checks seem to rule out the explanations pertaining to the unobserved quality differences between male and female teachers. Finally, we find that this negative effect in math not only seems to disappear but becomes (marginally) positive in the classrooms where the female teacher had a math or a math-related major in college/post-college. Thus, only female students taught by female teachers with *limited* math backgrounds appear to be adversely affected.

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, and utilize reading test scores as a robustness check—and provide *suggestive* evidence regarding the potential mechanisms for our results (discussed in detail in Section 3.3).

The remainder of the paper is organized as follows. Section 2 describes the Teach for America Program and the data. Section 3 discusses threats to identification, as well as the estimation strategy and results. Conclusions are presented in Section 4.

---

[8] While a number of recent studies based on nationally representative samples have shown that male and female students exhibit only very small differences in their math achievement through the end of high school (see for example, Shauman and Xie [2003]), they are not examining the differential effect of teacher gender on the gender gap in math test scores, nor are they examining the gender math achievement gap among the population of students where these gaps may be more pronounced, i.e., disadvantaged neighborhoods.

2. Teach for America Program and Data

2.1 Teach for America

Teach for America (TFA) is a non-profit 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 two-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.[9]

TFA corps members are placed in schools that are in some of the most disadvantaged neighborhoods in the United States. The challenge of being an efficient teacher in these environments without any or with very little previous teaching experience raised questions about the effectiveness of the program. Although each TFA corps member goes through a very competitive selection process and receives a short but very intense training program including significant support following training, critics argue that TFA teachers are still ill-prepared to teach in the challenging environments due to the lack of specialized education, training, and experience (Lackzo-Kerr and Berliner [2002], Pilcher and Steele [2005], Darling-Hammond [2005]).[10] Nevertheless, recent studies found that TFA teachers perform at least as well as, and in some cases better than, regular teachers (e.g., Decker et al. [2004], Kane et al. [2008], Antecol et al. [2012]).

---

[9] See http://www.teachforamerica.org.
[10] Other criticisms directed at the TFA program are the high attrition rate of TFA teachers and that more experienced and better trained teachers lose their jobs, in some cases to make room for TFA teachers (USA Today, 2009).

2.2 Data

We use data from the *Mathematica Policy Research, Inc* (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 2001-2002 academic year followed by full-scale evaluations in Chicago, Los Angeles, Houston, New Orleans and the Mississippi Delta in the next academic year (2002-2003). Each region had one school district participating in the experiment with the exception of 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. The final sample consisted of 17 schools, 100 teachers, and more than 1900 students in grades 1 through 5.

This data is 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 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 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 Section 3.1). 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 *pre-treatment* outcome variables); then at the end of

the academic year in which the study was conducted the students re-took math and reading tests based on the grade they just completed (which we call *post-treatment* outcome variables).

As previously noted the schools in our analysis were chosen to be broadly representative of the schools where TFA teachers taught at the time of the evaluation. 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 17 schools in the TFA study, the average rate of student eligibility for a free or a reduced-priced lunch was about 97 percent as opposed to 41 percent nationwide.

Although sample attrition is relatively small, we lose around 16 (18) percent 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 1,646 (1,606) students for the math (reading) test score sample from classes taught by 94 teachers.[11] To ensure the student composition was unaffected by the sample attrition we show that student characteristics, as well as teacher characteristics (if available), are similar for both the randomization sample and our estimation sample (see Columns 1 and 2 of Table 1, respectively).

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

---

[11] Our estimation samples (1,645 and 1,606 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.

coefficient on the teacher gender dummy is always small and has a large standard error irrespective of sample, i.e., the coefficient (standard error) for the whole randomization sample is 0.037 (0.034) and 0.015 (0.035) for the sample without missing pre-treatment math test scores.[12] Given sample attrition does not appear to be an issue, for the remainder of the discussion of descriptive statistics we focus on our estimation sample.

2.2.1 Student Characteristics

The data includes detailed information on student characteristics: type of class, gender, race and ethnicity, math test scores (pre- and post- treatment), reading test scores (pre- and post-treatment), and class size. 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 percent of the students (based on the math test score sample) were assigned to classrooms taught by a TFA teacher (see Column 2 of Table 1), and this assignment is not significantly different for male and female students.[13]

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.[14] Given the schools in our sample are disadvantaged schools the student body is predominantly non-white. Specifically, 66.3 (27.6) percent of the student body is black (Hispanic), while 6.1 percent is white. Furthermore, the average number of students per class is 25.1, the mean pre-treatment (baseline) normal curve equivalent (NCE) math test score is 30.3 while the mean pre-treatment (baseline) NCE reading test score is 25.6. Students in our sample are about one

---

[12] 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.

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

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

standard deviation below the national average (i.e., the NCE scale has a mean of 50 and standard deviation of 21 nationally).

2.2.2 Teacher Characteristics

The data also includes detailed information on teacher characteristics. Specifically, we have information on their gender, race/ethnicity, certification status, teaching experience, their major in college, and their major in their graduate degree if they have one. About one quarter of the students in our sample are taught by male teachers (see Table 1).[15] As previously noted, this allows us to examine the relative effectiveness of male and female teachers, and overcome a shortcoming of earlier research on teacher effectiveness in primary schools due to insufficient samples of male teachers.[16] Roughly 47 percent of the students have black teachers, 10 percent have Hispanic teachers, and 40 percent have white teachers.[17] For the remaining 3 percent, 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 non-regular certification (including emergency, temporary, initial, and other types of certifications). About 55 (45) percent of the students had teachers with regular (non-

---

[15] In our sample, about 33 percent of TFA teachers are male (i.e., 13 out 41 TFA teachers are male) as opposed to slightly less than 15 percent of control teachers (i.e., 8 out of 53 control teachers are male). Since the characteristics of TFA teachers are quite different than regular teachers in our sample (see Appendix 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.

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

[17] 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 Appendix Table A1).

regular) certification.[18] 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 Section 3.3 for a detailed discussion). As such we focus our attention on one particular set of college (undergraduate or graduate) majors: math or math-related (i.e., computer science and system analysis, engineering systems, pre-med, economics, and accounting) majors. Teachers who have a math or a math-related college major taught roughly 11.5 percent of the students in our sample.

3. Threats to Identification, Estimation Strategy and Results

3.1 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 OLS estimates. First, schools assign students to teachers non-randomly, even within subjects and grade levels (e.g., Clotfelter et al. [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 non-random 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

---

[18] 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 Appendix Table A1).

may be unobserved teacher traits that are correlated with student outcomes which again may bias the conventional OLS estimates; e.g., 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 non-random 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, non-random sorting of students to teachers of their parents' choice was avoided.[19] The second problem, on the other hand, is harder to deal with directly. As we 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 Column 1 of Table 2). As pre-treatment 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 pre-treatment math test scores which includes a control for pre-treatment math test scores (see

---

[19] Of course it would be interesting to show how important the non-random selection is for our results if we had comparable non-experimental data on primary schools. While we do not have direct evidence, we can point to the analysis by Dee (2007) who argues that the non-random 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 8th grade. Arguably, controlling for non-random 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 non-random selection from these other extensions.

Column 2 of Table 2), and an analogous regression for the sample of students with pre-treatment reading test scores (see Column 3 of Table 2). This reduces the sample size by 58 (121) observations for the sample of students with pre-treatment 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.[20] 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 mentioned in the preceding section, the data comes from very disadvantaged parts of the U.S. population such that almost 97 percent of the students are eligible for the free lunch program and that there are very few white students (slightly more than 6 percent). Therefore, we do not claim that our results can necessarily be generalized to the entire U.S. student population, but as previously noted they allow us to focus on a sample where the problems with the education system in the U.S. are well documented.

---

[20] 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 is generally statistically insignificant at conventional levels (the one exception is the coefficient on teacher gender in the black indicator regression is marginally significant, p-value 0.092).

3.2 Estimation Strategy and Results

In order to formally determine the effect of female teachers on student achievement in primary school we follow the existing economics literature and first estimate fully flexible regressions of the following form separately by student gender:

$$TS_{icb} = \beta_0 + \beta_1 FEMTEACH_{icb} + SC_{icb}^{i}\delta + TC_{icb}^{i}\gamma + \eta_b + \varepsilon_{icb} \tag{1}$$

where *TS* is subject-specific post-treatment student test scores for student *i* in class c and block *b* (i.e., classrooms in the same school and the same grade), and *FEMTEACH* is equal to 1 if the student's teacher is female and zero otherwise. *SC* is a vector of student characteristics (i.e., race/ethnicity, pre-treatment NCE math or NCE reading test scores, and class size) and *TC* is a vector of teacher characteristics (i.e., TFA indicator variable, race/ethnicity, years of teaching experience and its square, and type of certification). Finally, η represents block fixed effects, and ε is an error term with the standard properties.[21]

Table 3 presents our baseline results for post-treatment NCE math and NCE reading test scores by gender. Henceforth we refer to these as math and reading test scores. Following Dee (2008) 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.

We find that having a female teacher as opposed to a male teacher decreases math test scores for female students by about 4.7 percentage points (i.e., roughly 25 percent of a sample standard deviation). While being taught by a female teacher does also lead to lower reading test

---

[21] The standard errors from all regressions are clustered at the block-level.

scores for female students, the effect is much smaller and insignificant at conventional levels (column 6). The fact that female teachers appear to only lower the math achievement, but not the reading achievement, of female students suggests that the estimated effect of female teachers on math achievement is not biased by unobservable confounders. Furthermore, for male students, we do not observe any effect of having a female teacher either in math or reading test scores. Both coefficients are small, insignificant at conventional levels, and have opposite signs, i.e., the effect is positive (negative) for math (reading) scores. These results for male students further support the fact that the negative effect of female teachers on academic achievement is only limited to math scores of female students.

Overall, the results are suggestive that female teachers differentially influence the math achievement of male and female students in primary school and a similar effect is not found for reading.[22] In order to test whether the differential teacher gender effect we find in math for female and male students is statistically distinguishable, we estimate the following (more restricted) model:

$$
\begin{aligned}
TS_{icb} = \beta_0 &+ \beta_1 FEM_{icb} + \beta_2 FEMTEACH_{icb} + \beta_3 FEM_{icb} \times FEMTEACH_{icb} + \\
&SC_{icb}^{'}\delta + TC_{icb}^{'}\gamma + \eta_b + \varepsilon_{icb}
\end{aligned} \tag{2}
$$

where *FEM* is an indicator variable equal to one if the student is female and zero otherwise, and all other variables are as previously defined.

Columns 2 and 6 of Table 4 present the results from equation (2) for our two achievement measures, respectively. There are several noteworthy patterns. First, 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 percent of a sample standard deviation) on their math test scores.

---

[22]We find similar results if we limit our analysis to female teachers only as in Beilock et al. (2010), as well as female control teachers only. See detailed discussion in Section 3.3.

Second, 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 0.38). Finally, 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 (-3.37, which is a little less than 20 percent of a sample standard deviation) and statistically significant at the 10 percent level. Not surprisingly given the results in Table 3, we continue to find no significant effects of teacher gender on reading achievement of male and female students.[23]

Overall our results seem to confirm the results found in the educational psychology literature: female primary school teachers adversely influence the math outcomes of female students but not male students. The remainder of the paper examines the potential mechanisms that might help explain the influence of teacher gender on primary student outcomes.

3.3 Potential Mechanisms for the Math Achievement Outcomes of Female Students

How can we explain our findings in light of the fact that we have randomized data and that our estimates are not likely to be biased by unobserved confounders? 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 will disappear or become positive for the female students taught by female teachers with a strong math background.

---

[23] In order to ensure our results are not driven by blocks that do not include both male and female teachers, we re-run 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 available upon request. Ideally we would have re-estimated models that included controls for a teacher's math background (see Section 3.3) however the small number of teachers with a math background (i.e., 4 teachers) in these remaining 18 blocks did not afford us this option.

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 whether these differences in teaching styles differentially influence male and female students. Alternatively it could be that female teachers, with or without a strong math background, teach math as well as male teacher but female students respond better to a male teacher at this age. This seems less plausible however given 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 Table 5 and detailed discussion below). Finally, it could be that a female teacher with a strong math background is *less* likely to suffer from math anxiety. This is not only as a result of better math knowledge and ability but, perhaps, also indicates 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) found 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.[24,25] Moreover, they show that

---

[24] 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. 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.).

this negative effect works through female 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.[26,27]

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

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

where *MATHREL* is equal to one if the classroom was taught by a teacher with a math or a math-related (i.e., computer science and system analysis, engineering systems, pre-med, economics, and accounting) undergraduate or graduate major and zero otherwise, and all other variables are as previously defined. We must note that one should be cautious in interpreting the results based on these regressions since we only have 12 teachers with a math or a math-related major in our sample (9 female teachers with 143 students and 3 male teachers with 46 students).

The results are presented in Columns 4 and 8 of Table 4. For math test scores, the relative difference in the test scores between female students assigned to female versus male

---

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.

[25] Math anxiety among primary school teachers, of which females constitute about 90 percent, 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 (Bush [1989], Tobias [1998]) however they do not discuss different effects for male and female students.

[26] 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.

[27] This 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 s/he has the potential to confirm a negative stereotype about the social group the person belongs to.

teachers and male students with female versus male teachers *with* a math or a math-related major $(\beta_3 + \beta_7)$ is a very large positive number (4.745) and almost significant (with the p-value of 0.104) 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 math or a math-related major (β_3) is very large and negative (-4.393) and significant at the 5 percent level. Finally, the difference-in-difference-in-differences estimator (β_7) is very large and positive (9.138) and significant at the 1 percent level.

Although very large magnitudes for coefficient estimates raise some questions about their reliability, potentially because of the very small sample of male teachers who have math or math-related backgrounds, these results *tentatively* suggest that the negative effect of being assigned to a female teacher on math test scores for female students results from female teachers *without* a math or a math-related major. When we look specifically at teachers with math and math-related majors, this negative effect not only seems to disappear but it becomes positive, although only marginally significant. This appears to be consistent with the results found in Carrell et al. (2010), who, using experimental data, find that highly selective female professors, who by definition have extensive training in math, positively influence female student math achievement outcomes at the U.S. Air Force Academy.[28] 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.[29]

---

[28] While Hoffman and Oreopoulos (2009) also find a small positive effect at a selective institution of higher education, their data is non-experimental.

[29] 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

In order to ensure that our results are not driven by the small number of male teachers with a math or a math-related major, following Beilock et al. (2010), we estimate the following two specifications using female teachers only:

$$TS_{icb} = \lambda_0 + \lambda_1 FEM_{icb} + SC'_{icb}\varpi + TC'_{icb}\vartheta + \kappa_b + u_{icb} \tag{4}$$

and

$$\begin{aligned} TS_{icb} = \lambda_0 + \lambda_1 FEM_{icb} + \lambda_2 MATHREL_{icb} + \lambda_3 FEM_{icb} \times MATHREL_{icb} \\ + SC'_{icb}\varpi + TC'_{icb}\vartheta + \kappa_b + u_{icb}. \end{aligned} \tag{5}$$

where λ_1 captures the gender gap in test scores for those students taught by a female teacher in equation (4) and by a female teacher *without* a strong math background in equation (5). λ_3 in equation (5) captures the gender gap for those students taught by female teachers *with* a strong math background versus those taught by female teachers *without* a strong math background. Since we control for pre-treatment test scores, these coefficients will capture the gender gap in test scores generated within our sample period (an academic year).

Table 5 presents the results from equations (4) and (5) for our two test score measures. Female students taught by female teachers have math test scores that are significantly lower (-1.798) than their male counterparts (Column 2), that is, lower by roughly 9 percent of a standard deviation. Moreover, the math test score gap remains similar (-2.161) if the classes are taught by female teachers with no math background (column 4). The gender gap, however, disappears and the coefficient estimate becomes positive (1.086), although statistically insignificant, for students taught by female teachers with a strong math background. Finally, the difference between these two effects is large (3.247) and statistically significant at 5 percent

---

scores of female students in primary school. Moreover, the remainder of the analysis excludes male teachers and our results continue to support these findings.

level. In other words, female students in classes taught by female teachers *with* a math or a math-related major score roughly 3 percentage points higher on their math test scores than female students taught by female teachers *without* a math background. In terms of reading test scores, we find that none of the coefficients of interest are statistically significant. These results are consistent with our previous findings suggesting that the small number of male teachers with a math or a math-related major do not appear to be biasing our results.

We also check if including TFA teachers is biasing our results by re-estimating equations (4) and (5) for control teachers only. The results are qualitatively the same for both math and reading test scores, although in some cases we lose precision (see Table 6). This is likely largely an artifact of the small number of female teachers with a math background being further disaggregated by teacher status, i.e., only 6 (out of 53) female control teachers have a math or a math-related degree. Ideally we would have re-estimated models that included both male and female teachers for control teachers only. However, the small number of male control teachers combined with the small number of male teachers with a math background did not afford us this option. Moreover, it is the very fact that we have TFA teachers that allowed us to examine the relative effectiveness of male and female teachers which, as previously stated, was a shortcoming of earlier research on teacher effectiveness in primary schools due to insufficient samples of male teachers.

Overall our findings show that female students fare worse on their math test scores if they were taught by a female teacher *without* a strong math background. This is true whether we compare female teachers to male teachers or we look at female teachers in isolation. Specifically, this negative effect is 9-10 percent of a standard deviation of math test scores when our specification includes students taught by female teachers only and is roughly twice as large

when our specification includes students taught by both female and male teachers. Taking the 2004 National Assessment of Educational Progress (NAEP) math test score averages at age 9 as our benchmark, our lower bound estimate is about the same size as the raw gender test score gap or the standardized gender gap in math test scores at age 13 on the 1999 NAEP (Dee [2007]). Our lower bound estimate is also in-line with Dee (2007) who finds the negative effect of female teachers on the math test scores of eighth grade students is roughly 6-8 percent of a standard deviation. However, as we pointed out before, he also finds this effect for male students and after conducting several robustness checks he concludes that his findings on math achievement are largely due to the non-random assignment of female teachers to classrooms with low performing students in math.

While our upper bound estimate at first glance may seem high relative to the existing economics literature, 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. Specifically, we not only focus on students in primary schools, more than 90 percent of the students in our sample 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 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, 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 non-randomized data.

In addition, the effect of teacher gender on student achievement presented in the existing economics literature for higher levels of education (particularly post-secondary education) may indeed to be more closely aligned with the positive (although not always significant) 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 selective 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.

4. Conclusion

Are female students adversely affected by having a female teacher? The evidence in the economics literature—which focuses on higher levels of education—is mixed. These studies either find 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]) or *no* effect on female achievement

outcomes (e.g., Canes and Rosen [1995], Ehrenberg et al. [1995], Neumark and Gardecki [1998]) in middle school/high school or college/post-college.

While the impact of teacher gender on the outcomes of students in primary school, to the best of our knowledge, has received no attention to date in the economics literature, the same is not true for the educational psychology literature. Specifically, Beilock et al. (2010) finds that having a female teacher has a *negative* effect on female student math achievement in primary school.[30] Moreover, they show that this negative effect is an artifact of the *math anxiety hypothesis*—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 male students or female students without such a belief.

This paper further analyzes the effect of teacher gender on primary student outcomes using evidence from a well-executed randomized experiment. Our unique data allows us to avoid the issue of non-random assignment of teachers, as well as affords us a large sample of primary schools, students, teachers, and states. Furthermore, our data comes 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 provides 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.

---

[30] Earlier studies also find a negative effect of math anxiety on teaching performance in math classes (i.e., Bush [1989], Tobias [1998]) however they do not discuss different effects for 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.

In an attempt to further understand the potential mechanisms behind our results, we draw on the educational psychology literature and *tentatively* show that female students' internalization of female teachers' math anxiety may play a role under the assumption that having a math or a math-related college/post-college major likely reduces or eliminates math anxiety among female primary school teachers. In other words, we posit math anxiety may be reduced either because female teachers have more extensive training in math and/or they may not hold stereotypical beliefs about gender ability differences. Specifically, the negative effect of having a female teacher on math test scores of female students not only seems to disappear but becomes (marginally) positive 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 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

24

anxiety among female teachers and students' beliefs of gender stereotypes, 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 (Hembree [1990], Emenaker [1996], Vinson [2001] among many others). 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, with available randomized data, should address this important issue in further detail.

References

Altonji, Joseph G. and Rebecca Blank. 1999. "Race and Gender in the Labor Market," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics Volume*, 3c Elsevier Science B.V. : 3144-3259.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2012. "The Effect of the Teach for America Program on Student Test Scores: A Distributional Analysis using Randomized Data," Mimeo, Claremont McKenna College.

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(5)*, 1060-1063.

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: 152–157.

Bursal, Murat. and Lynda Paznokas. 2006. "Mathematics Anxiety and Preservice Elementary Teachers' Confidence to Teach Mathematics and Science," *School Science and Mathematics*, 106(4): 173-179.

Bush, William S. 1989. "Mathematics Anxiety in Upper Elementary School Teachers" *School Science and Mathematics*, 89: 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: 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(3): 1101-1144.

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(4): 778-820.

Darling-Hammond, Linda, Deborah J. Holtzman, Su Jin Gatlin, and Julian V. Heilig. 2005. "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness," *Education Policy Analysis Archives*, 13(42).

Decker, Peter T., Daniel P. Mayer, and Steven Glazerman. 2004. "The Effects of Teach for America on Students: Findings from a National Evaluation." Mathematica Policy Research Report 8792-8750, New York.

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

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95(2): 158-65.

Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources* 42(3): 528-554.

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: 830–847.

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: 547–561.

Emenaker, Charles. 1996. "A Problem-Solving Based Mathematics Course and Elementary Teachers' Beliefs," *School Science and Mathematics,* 96(2): 75-84.

Gresham, Gina. 2007. "A Study of Mathematics Anxiety in Pre-Service Teachers," *Early Childhood Education Journal,* 35:181–188.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Using Value-Added Measures of Teacher Quality," *American Economic Review*, 100(2): 267-271.

Hembree, Ray. 1990. "The Nature, Effects, and Relief of Mathematics Anxiety," *Journal for Research in Mathematics Education,* 21(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(2): 479-494.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review*, 27: 615-631.

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(3):587-613.

Kaplan, Karen. 2010. "Female Teacher May Pass on Math Anxiety to Girls, Study Finds," *Los Angeles Times*, January 26.

Laczko-Kerr, Ildiko and David C. Berliner. 2002. The Effectiveness of "Teach for America" and Other Under-Certified Teachers on Student Academic Achievement: A Case of Harmful Public Policy*, Education Policy Analysis Archives*, 10(37).

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.

National Science Foundation. 2006. "Science and Engineering Degrees: 1966–2004," Manuscript NSF 07-307, National Science Foundation, Division of Science Resources Statistics.

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: 220–246.

Pilcher, Lorene C., and Donald C. Steele. 2005. "Teach for America and Regularly Certified Teachers: Teacher Efficacy, Teaching Concerns, Career Aspirations, and Teaching Effectiveness," In *Research on Alternative and Non-traditional Education*, p. 123-142 Lanham, Maryland: The Association of Teacher Educators.

Pope, Devin G., and Justin R. Sydnor. 2010. "A New Perspective on Stereotypical Gender Differences in Test Scores," *Journal of Economic Perspectives*, 24: 95–108.

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

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay and Student Achievement," *Quarterly Journal of Economics*, 25(1): 175-214.

Shauman, Kimberlee A. and Yu Xie. 2003. *Women in Science: Career Processes and Outcomes.* Harvard University Press, Cambridge, Massachusetts.

Steele, Claude M. 1997. "A Threat in the Air—How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, 52(6): 613–29.

Steele, Jennifer. 2003. "Children's Gender Stereotypes about Math: The Role of Stereotype Stratification," *Journal of Applied Social Psychology*, 33: 2587–2606.

Tobias, Sheila. 1998. "Anxiety and Mathematics," *Harvard Educational Review*, 50: 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*(2): 89-94.

Weinberger, Catherine J. 1999. "Mathematical College Majors and the Gender Gap in Wages," *Industrial Relations*, 38: 407–413.

Weinberger, Catherine J. 2001. "Is Teaching More Girls More Math the Key to Higher Wages?" in *Squaring Up: Policy Strategies to Raise Women's Incomes in the U.S.*, Mary C. King, ed. (Ann Arbor: The University of Michigan Press).

Table 1 – Descriptive Statistics by Sample
(Means and Standard Errors)

	Randomization Sample	Estimation Sample
	(1)	(2)
Student Characteristics		
Female	0.489	0.494
	(0.011)	(0.012)
White	0.069	0.061
	(0.006)	(0.006)
Black	0.668	0.663
	(0.011)	(0.012)
Hispanic	0.263	0.276
	(0.010)	(0.011)
Class size (number of students)	25.152	25.146
	(0.129)	(0.138)
Pre-Treatment Math (average NCE) +	29.692	30.272
	(0.4149)	(0.461)
Pre-Treatment Reading (average NCE) +	25.064	25.600
	(0.494)	(0.519)
Teacher Characteristics		
Female	0.762	0.756
	(0.010)	(0.011)
TFA	0.440	0.440
	(0.010)	(0.012)
Certified	-	0.546
		(0.012)
Teacher Experience (in years)	-	6.134
		(0.196)
White	-	0.401
		(0.012)
Black	-	0.474
		(0.012)
Hispanic	-	0.102
		(0.008)
Race Missing	-	
Math or Math Related Major	-	0.115
		(0.008)
Number of observations	1920	1646

Notes: Neither sample includes 49 students for whom we do not have information on the gender of their teachers. +Pre-treatment test scores are based on slightly smaller samples for the randomized sample due to non-reporting; 1862 students with pre-treatment Math test scores and 1799 students with pre-treatment Reading test scores.

Table 2 - Test for Randomization
(Coefficients and Standard Errors)

	Dependent Variable: Female Teacher Dummy		
	All Students	All Students with Pre-Treatment Math Test Scores	All Students with Pre-Treatment Reading Test Scores
	(1)	(2)	(3)
Female Student	-0.010	-0.012	-0.007
	(0.008)	(0.009)	(0.009)
Class size	0.005	-0.004	-0.010
	(0.019)	(0.021)	(0.026)
Black Student	-0.137	-0.162	-0.164
	(0.119)	(0.135)	(0.139)
Hispanic Student	0.103	0.080	0.080
	(0.135)	(0.146)	(0.147)
Pre-Treatment Math	-	0.001	-
		(0.001)	
Pre-Treatment Reading	-	-	-0.001
			(0.0001)
Block Fixed Effects	Yes	Yes	Yes
Number of Observations	1,920	1862	1799
R-squared	0.420	0.397	0.411

 Notes: We control for a TFA indicator variable in all regressions given TFA status is the treatment variable in the original experiment. Standard errors are clustered at the block level.
 *** Significant at the 1 percent level; ** Significant at the 5 percent significance level; *Significant at the 10 percent significance level.

Table 3 – Determinants of Math and Reading Test Scores by Student Gender
(Coefficients and Standard Errors)

	Math					Reading			
	Female	Female	Male	Male		Female	Female	Male	Male
Female Teacher	-5.598***	-4.683***	0.655	0.381		-1.230	-0.651	-0.479	-0.612
	(1.465)	(1.574)	(1.850)	(2.312)		(1.107)	(1.418)	(1.521)	(1.480)
Teacher Characteristics	No	Yes	No	Yes		No	Yes	No	Yes
Student Characteristics	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Block Fixed Effcts	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Number of Observations	813	813	833	833		796	796	810	810
R-squared	0.586	0.589	0.549	0.550		0.745	0.749	0.661	0.666

Notes: Student characteristics include race/ethnicity, pre-treatment NCE math or NCE reading test scores, and class size. Teacher characteristics include a TFA indicator variable, race/ethnicity, years of teaching experience and its square, and type of certification. Standard errors are clustered at the block level.
*** Significant at the 1 percent level; ** Significant at the 5 percent significance level; *Significant at the 10 percent significance level.

Table 4 – Determinants of Math and Reading Test Scores Pooled by Student Gender
(Coefficients and Standard Errors)

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female (β_1)	1.528	1.482	2.187	2.162	0.975	1.042	1.220	1.278
	(1.662)	(1.653)	(1.606)	(1.604)	(1.510)	(1.516)	(1.625)	(1.635)
Female Teacher (β_2)	-0.673	-0.397	0.153	0.377	-0.816	-0.819	-0.314	-0.466
	(1.726)	(2.069)	(1.943)	(2.229)	(1.295)	(1.185)	(1.481)	(1.387)
Female*Female Teacher (β_3)	-3.432*	-3.374*	-4.419**	-4.393**	-0.148	-0.258	-0.519	-0.614
	(1.882)	(1.871)	(1.905)	(1.899)	(1.485)	(1.493)	(1.651)	(1.671)
Math Major (β_4)	-	-	2.195	2.490	-	-	1.638	-0.738
			(4.776)	(5.507)			(1.296)	(1.660)
Femal Teacher*Math Major(β_5)	-	-	-6.107	-6.241	-	-	-3.754	-2.460
			(5.325)	(6.195)			(2.463)	(2.701)
Female*Math Major (β_6)	-	-	-5.950**	-5.970**	-	-	-2.973	-2.989
			(2.659)	(2.658)			(2.919)	(2.934)
Female*Female Teacher*Math Major (β_7)	-	-	9.112***	9.138***	-	-	4.203	4.220
			(3.173)	(3.210)			(4.671)	(4.706)
$\beta_1 + \beta_3$	-1.905**	-1.892**	-	-	0.827	0.784	-	-
	(0.742)	(0.740)			(0.801)	(0.801)		
$\beta_3 + \beta_7$	-	-	4.693*	4.745	-	-	3.684	3.606
			(2.796)	(2.841)			(3.919)	(3.933)
Teacher Characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Student Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Block Fixed Effcts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	1,646	1646	1,646	1,646	1,606	1,606	1,606	1,606
R-squared	0.554	0.554	0.555	0.556	0.695	0.697	0.696	0.698

Notes: See notes to Table 3.

Table 5 – Determinants of Math and Reading Test Scores Pooled by Student Gender for Female Teachers Only
(Coefficients and Standard Errors)

	Math					Reading			
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
Female (λ_1)	-1.756**	-1.798**	-2.082**	-2.161**		0.835	0.814	0.698	0.667
	(0.780)	(0.787)	(0.844)	(0.860)		(0.823)	(0.824)	(0.791)	(0.803)
Math Major (λ_2)	-	-	-4.597	-3.742		-	-	-2.009	-2.795
			(3.051)	(3.270)				(2.628)	(2.795)
Female*Math Major (λ_3)	-	-	3.135**	3.247**		-	-	1.281	1.297
			(1.458)	(1.512)				(2.912)	(2.938)
$\lambda_1 + \lambda_3$	-	-	1.053	1.086		-	-	1.979	1.964
			(1.311)	(1.330)				(2.898)	(2.908)
Teacher Characteristics	No	Yes	No	Yes		No	Yes	No	Yes
Student Characteristics	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Block Fixed Effcts	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Number of Observations	1245	1245	1245	1245		1218	1218	1218	1218
R-squared	0.546	0.552	0.548	0.553		0.743	0.745	0.743	0.745

Notes: See notes to Table 3.

Table 6 – Determinants of Math and Reading Test Scores Pooled by Student
Gender for Female Control Teachers Only
(Coefficients and Standard Errors)

	Math		Reading	
	(1)	(2)	(3)	(4)
Female (λ_1)	-2.457**	-2.791**	0.377	0.160
	(1.085)	(1.104)	(1.000)	(0.880)
Math Major ($\lambda2$)	-	-4.186	-	-2.425
		(3.675)		(4.369)
Female*Math Major (λ_3)	-	3.874	-	2.431
		(2.301)		(4.722)
$\lambda_1 + \lambda_3$	-	1.083	-	2.591
		(2.163)		(4.749)
Teacher Characteristics	Yes	Yes	Yes	Yes
Student Characteristics	Yes	Yes	Yes	Yes
Block Fixed Effcts	Yes	Yes	Yes	Yes
Number of Observations	771	771	757	757
R-squared	0.556	0.517	0.748	0.748

Notes: See notes to Table 3.

Appendix

Table A1 - Teacher Characteristics by Treatment Status and Gender

	Overall				Control				TFA		
	Total	Male	Female		Total	Male	Female		Total	Male	Female
Race/Ethnicity (Percentage)											
African American	47.9	28.6	53.4		69.8	62.5	71.1		19.5	7.7	25
White	38.2	57.1	34.2*		11.4	0	15.6		73.2	92.3	64.3*
Hispanic	10.7	14.3	9.6		15.1	37.5	11.1*		4.9	0	7.1
Race Missing	3.2	0	2.6		1.9	0	2.2		2.4	0	3.6
Certification (Percentage)											
Regular	55.3	28.6	63***		64.2	12.5	73.3***		43.9	38.5	46.4
Non-Regular	44.7	71.4	37***		35.8	77.5	16.7**		56.1	61.5	53.6
Average Years of Teaching Experience	6.3	4.1	6.9		9.7	7.8	10		2	1.8	2
Number of Teachers	94	21	73		53	8	45		41	13	28

*** Significantly different than males at the 1 percent level; ** Significantly different than males at the 5 percent significance level; *Significantly different than males at the 10 percent significance level.