TEACHERS AND THE GENDER GAPS
IN STUDENT ACHIEVEMENT

Thomas S. Dee

Teachers and the Gender Gaps in Student Achievement
Thomas S. Dee
NBER Working Paper No. 11660
September 2005
JEL No. I2

ABSTRACT

In the United States, girls outperform boys in measures of reading achievement while generally underperforming in science and mathematics. One major class of explanations for these gaps involves the gender-based interactions between students and teachers (e.g., role-model and Pygmalion effects). However, the evidence on whether these interactions actually matter is limited and contradictory. In this study, I present new empirical evidence on whether assignment to a same-gender teacher influences student achievement, teacher perceptions of student performance, and student engagement. This study's identification strategy exploits a unique "matched pairs" feature of a major longitudinal survey. Within-student comparisons based on these data indicate that assignment to a same-gender teacher significantly improves the achievement of both girls and boys as well as teacher perceptions of student performance and student engagement with the teacher's subject. For example, assignment to a female science teacher increases the likelihood that a girl views science as useful for her future. However, because the middle-school teachers in most academic subjects are female, these results also suggest that the gender dynamics between teachers and students at this level amplify boys' large underperformance in reading while attenuating the more modest underperformance of girls in math and science.

Thomas S. Dee
Department of Economics
Swarthmore College
Swarthmore, PA 19081
and NBER
dee@swarthmore.edu

1. INTRODUCTION

Society's fundamental interest in fairness and equal opportunity continues to motivate highly contentious debates over the root causes of gender differences in educational outcomes. Much of this heated discussion, both in popular and academic settings, has focused on assessing the relative contributions of biological and environmental determinants (i.e., "nature versus nurture").[1] However, there are also pointed disagreements among studies that stress the role of environmental influences. In particular, the so-called "gender wars" have recently offered sharply contrasting images of how interactions with teachers may shape the relative cognitive development and intellectual engagement of boys and girls (AAUW 1992, Sommers 2000, Sadker 2002).[2]

This recent discussion has centered on controversial claims that teachers consistently privilege boys over girls (e.g., with more positive feedback and helpful questions).[3] However, the literature in this area has also focused whether these classroom dynamics differ when a student and teacher share the same gender. Assignment to a same-gender teacher could be educationally relevant for a number of reasons. For example, it could influence student engagement or behavior through role-model effects and stereotype threat. Furthermore, same-gender teachers may also communicate different (and self-fulfilling) expectations to the boys and girls in their classrooms (i.e., Pygmalion effects). Prior studies have examined the empirical relevance of these dynamics by assessing the reduced-form impact of assignment to a same-gender teacher on educational outcomes (e.g., Bettinger and Long 2005, Canes and Rosen 1995, Ehrenberg,

---

[1] For example, the recent remarks on this topic by Larry Summers have drawn considerable public attention to this topic (Dillon 2005). Similarly, a more recent comment by First Lady Laura Bush that female teachers "are great in language arts and not so great in math" (Abcarian 2005) also drew attention.

[2] However, concerns over whether school practices favor one gender over another are far from new. For example, nearly a century ago, Ayres (1909) expressed concern over the relative reading achievement of boys in the elementary grades.

[3] See AAUW (1992) and Sadker and Sadker (1994) for evidence in support of these claims and Saltzman (1994), Kleinfeld (1998), Lewin (1998), and Sommers (2000) for critical appraisals.

Goldhaber and Brewer 1995, Rothstein 1995, Neumark and Gardecki 1998, Nixon and Robinson 1999, and Robst, Keil and Russo 1998). However, the results of these studies have been mixed.

This study presents new evidence on whether assignment to a same-gender teacher influences educational outcomes. And it uses these evaluation results to discuss how the gender interactions between students and teachers may shape the early evolution of gender gaps across academic subjects. The evidence presented here makes three distinct contributions to the extant literature. First, this study focuses on a nationally representative sample of middle-school students (i.e., 8th graders) instead of students at secondary or postsecondary levels. This distinction may be particularly relevant because early adolescence corresponds more closely with the age at which the gender gaps in educational achievement become particularly pronounced (e.g., Table 1).

Second, in addition to test scores, the educational outcomes examined in this study include teacher perceptions of a student's performance and student perceptions of the subject taught by a particular teacher. These subjective outcomes are useful simply because they are educationally important outcomes that provide a complement to the results based on achievement scores. However, measures of student engagement with specific academic subjects may also be particularly important as precursors of the subsequent gender gaps in curricula and occupations. For example, college females are underrepresented in fields like computer science and engineering to an extent that may have more to do with their confidence and interest in math and science than with the relatively modest gender gaps in their prior math and science achievement (e.g., Cavanagh 2005).

Third and perhaps most important, this study adopts an identification strategy that exploits a unique "matched pairs" feature of the National Education Longitudinal Study of 1988

(NELS:88). In its base year, NELS:88 surveyed a nationally representative cross section of nearly 25,000 8th graders. However, NELS:88 also surveyed *two* of each student's academic-subject teachers. These surveys elicited information on each teacher's background and on how each teacher viewed the performance of the sampled student. Furthermore, NELS:88 also fielded student tests in the subjects taught by these teachers and gathered self-reported information from students on their attitudes towards the teacher's academic subject. This unique "cluster sample" (e.g., Wooldridge 2002, page 328) implies that the effects of a same-gender teacher can be identified in models that control for the influence of unobserved student traits, which may have biased the conventional cross-sectional evaluations. This analysis is similar to that used for data on monozygotic twin pairs (e.g., Ashenfelter and Krueger 1994, Ashenfelter and Rouse 1998, and Rouse 1999). But, in this context, the paired observations are of the same student observed contemporaneously in two different academic subjects.

The next section briefly characterizes the gender gaps in educational outcomes and discusses the literature on the sources of these differences. Sections 3 and 4 introduce the data and specifications used to assess the reduced-form effects of gender interactions. Sections 5 and 6 discuss the basic test score results and a variety of specification checks that assess the internal validity of these inferences.[4] Section 7 presents the results for the outcome measures based on teacher and student perceptions. The final section concludes with a discussion of the effect sizes documented here and with a brief assessment how the gender dynamics between students and teachers may influence the gender gaps in outcomes in light of the gender distribution of teachers across subjects.

---

[4] Though these matched-pairs comparisons control for subject fixed effects and student fixed effects that are constant across subjects, there are several possible threats to the internal validity of these inferences (e.g., non-random sorting by students with subject-specific propensities for achievement and unobserved teacher traits).

2. GENDER GAPS AND TEACHERS

Recent data from the recent Early Childhood Longitudinal Study indicate that, on entering kindergarten, boys and girls perform similarly on tests of general knowledge, reading and mathematics (Freeman 2004). However, by the spring of the 3rd grade, boys have slightly higher mathematics scores and lower reading scores. These subject-specific gender gaps appear to expand as students advance through the elementary and secondary grades. Table 1 presents average scores by gender, subject and student age on the 1999 National Assessment of Educational Progress (NAEP). Among 9 year-olds, boys have higher mathematics and science scores but a lower average score in reading. The relative underperformance of boys in reading is particularly large (0.16 of a standard deviation). And, among 9 year olds, it is the only difference that is statistically significant.

However, the NAEP data indicate that, over the next four years, each of these subject-specific achievement gaps increases. Specifically, as we move from 9 to 13 year-olds, the standardized gender gaps roughly double in science and reading and become statistically distinguishable from zero (Table 1). The standardized gender gap in math scores also increases (by roughly two-thirds) and becomes weakly significant. As we move from 13 to 17 year olds, the gender gap in science achievement continues to expand. But, interestingly, there is very little growth in the math and reading gender gaps among teenagers.[5]

The sizes of the gender gaps in test scores as students complete their secondary schooling are not trivial. For example, the underperformance of 17 year old boys in reading (i.e., 0.3 standard deviations) is nearly half the size of the corresponding black-white differential, a topic

---

[5] One notable caveat is that changes in average NAEP scores by gender between the ages of 13 and 17 may be biased by the fact that boys are more likely than girls to drop out of high school. However, this does not appear to be too empirically relevant in these circumstances. Specifically, this sort of differential attrition would imply that the more modest growth in the math and science gender gap is overstated while the growth in the reading gender gap is understated.

which has also commanded considerable attention. Another way to benchmark the relatively poor reading proficiency of boys is to note that it is roughly equivalent to 1.5 years of schooling (Riordan 1999). The underperformance of 17 year old girls in science and math (i.e., 0.2 and 0.1 standard deviations, respectively) is more modest but still qualitatively large (e.g., 20 and 10 percent of the respective black-white gaps).

The gender differences in these average test scores, of course, identify only some of the ways that the educational outcomes of boys and girls differ. For example, Hedges and Nowell (1995) find that the variance in male test scores is consistently larger than that of females, implying that, for several types of outcomes, boys are overrepresented among both high and low performers. Boys are also substantially more likely than girls to repeat a grade (Freeman 2004). And boys are now increasingly less likely than girls both to attend college and to persist in attaining a degree (e.g., Jacob 2002).[6] However, female college students continue to be underrepresented in certain technical fields like engineering or computer science (Freeman 2004).

The sources of the gender differences in educational outcomes have been the subject of considerable study and debate. One particularly contentious issue involves the possible role played by biological differences between males and females. Tests of general intelligence suggest that there are no overall differences between males and females. However, there do appear to be large gender differences with respect to average scores on specific cognitive tasks. For example, males outperform females at visual-spatial tasks, which are thought to complement mathematical problem-solving, whiles females excel at certain verbal tasks. A task-force report sponsored by the American Psychological Association in response to the publication of "The

---

[6] The fact that boys lag behind girls with respect to a variety of important educational outcomes has been called the "silent gender gap" (Riordan 1999).

Bell Curve" suggested that biological factors do contribute to these gender differences (Neisser et al. 1996). In particular, there are differences in male and female brain structures and in exposure to sex hormones that appear to influence the gender-specific skill advantages.[7] However, Neisser et al. (1996) note that these biological differences interact with environmental factors that appear soon after birth. Furthermore, discussions of this evidence frequently emphasize that there is "substantial overlap in the distribution of male and female scores" (e.g., Coley 2001).

The gender dynamics in classrooms are also frequently portrayed as an important "environmental" source of the gender differences in educational outcomes (e.g., AAUW 1992, Sommers 2000). There are a number of structural explanations for why assignment to a same-gender teacher, in particular, might influence the educational experiences of boys and girls. And understanding the distinctions among these theoretical explanations is an important antecedent to designing well-targeted policy interventions.

One broad hypothesis is that male and female teachers have unique biases with respect to how they engage boys and girls in the classroom. For example, there is controversial evidence based on classroom observations that teachers are more likely to offer praise and remediation in response to comments by boys but mere acknowledgement in response to comments by girls (e.g., AAUW 1992, Sadker and Sadker 1994, Saltzman 1994, Kleinfeld 1998, Lewin 1998, and Sommers 2000). Similarly, cognitive process theories (e.g., Jones and Dindia 2004) suggest that teachers may subtly communicate that they have different academic expectations of boys and girls. And these biased expectations become self-fulfilling when students respond to them (i.e., a Pygmalion effect).

---

[7] For more recent discussions of this evidence, see Kimura (1999), Halpern (2000), Lippa (2002), and Cahill (2005).

The available evidence on the extent to which male and female teachers share any particular bias in how they interact with girls or boys is more limited and contradictory. For example, in a recent literature review, Jones and Dindia (2004) cite several small-scale studies in support of this claim and conclude that a teacher's gender is "the most obvious factor that seems to shape sex equity in the classroom." However, these studies focused exclusively on post-secondary settings. In an earlier review of research spanning different grade levels, Brophy (1985, page 137) concludes that "teachers do not systematically discriminate against students of the opposite sex."

A second class of explanations for the educational relevance of a teacher's gender involves how students respond to a teacher's gender and not how that teacher actually behaves. For example, the potential existence of a "role model" effect implies that a student will have improved intellectual engagement, conduct, and academic performance when assigned to a same-gender teacher. A recent variation on this argument is the phenomenon known as "stereotype threat." This refers to a situation where student performance suffers when they fear being viewed through the lens of a negative stereotype threat. A recent experimental study by Spencer, Steele, and Quinn (1999) suggests that stereotype threat does apply to female performance in math. Specifically, they found that female subjects underperformed on a math test when told that the test produces gender differences but did not when told the opposite.

This study does not attempt to distinguish among these structural explanations but instead provides reduced-form evidence on the educational consequences of assignment to a same-gender teacher. Interestingly, most prior evidence on this question has focused on postsecondary and graduate settings (e.g., Canes and Rosen, 1995, Rothstein 1995, Neumark and Gardecki 1998, Robst, Keil and Russo 1998, and Bettinger and Long, 2005). The conclusions from these

studies are quite mixed as are the ones from the fewer studies that have examined the effect of a teacher's gender in high-school settings.

For example, Nixon and Robinson (1999), using data from the National Longitudinal Survey of Youth (NLSY), found that females attending high schools with a higher proportion of female faculty had higher levels of educational attainment. They also found no association between the presence of female faculty and the educational attainment of male students. In contrast, using cross-sectional data on 10th graders participating in the National Education Longitudinal Study of 1988 (NELS:88), Ehrenberg, Goldhaber and Brewer (1995) found that a teacher's gender was not associated with the achievement gains of girls (or boys). However, they did find that white female students were more favorably evaluated by white female teachers in math and science. A recent study by Lavy (2004), based on blind and non-blind test-score data from Israeli schools, found evidence that public high school teachers discriminated against male students and that these effects varied by the teacher's gender and subject.

As noted earlier, this study contributes to this literature in three ways. First, this study focuses on younger students (i.e., 8th graders) who are closer to the age when the gender gaps in achievement grow rapidly. Second, unlike most studies, this research focuses on several different student outcomes (i.e., test scores, teacher perceptions of student performance, and student perception of a particular academic subject), which are described in the next section. Third, this study also adopts a simple panel-based identification strategy that eliminates some (but not all) of the potential biases that could compromise the conventional cross-sectional evidence. Specifically, a possible problem with cross-sectional evaluations in this context is that they may be biased by the non-random assignment of students to teachers. For example, the prior evidence that females have better outcomes when with female teachers could occur if females with an

unobserved propensity for achievement are more likely to be matched with female teachers. Similarly, if boys with a lower propensity for achievement are more likely to be assigned to male teachers, the estimated benefits of a male teacher would be biased downwards. This study addresses such concerns by evaluating the effects of a teacher's gender on student outcomes in specifications that effectively condition on student fixed effects. This approach is feasible because of a unique, "matched-pairs" feature of a major longitudinal study, which is described in the next section.

3. NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988 (NELS:88)

The National Education Longitudinal Study of 1988 (NELS:88) is a nationally representative, longitudinal study that began in 1988 with a sample of 24,599 8th grade students from 1,052 public and private schools (Ingels et al. 1990). NELS:88 relied on a two-stage sampling design. In the first stage, schools, which were the primary sampling unit, were selected with probabilities proportional to their 8th grade enrollment. Approximately 26 students were then randomly chosen within each participating school.

NELS:88 also fielded questionnaires to the teachers responsible for teaching each of the selected students in two of four academic subjects: mathematics, science, reading, and social studies. The two surveyed teachers were chosen by randomly assigning each sampled school to one of four subject-area groupings: mathematics/reading, mathematics/social studies, science/reading, and science/social studies. The teacher survey solicited information about the teacher's background and about how the teacher viewed the behavior and performance of the sampled student. The student component of NELS:88 also collected test scores in the subject taught by that teacher as well as the student's perception of the subject taught by that teacher.

This arrangement implies that we observe a variety of student-level educational outcomes (i.e., test scores, teacher perceptions, and student perceptions). But, more important, it implies that each student-level outcome is observed twice. That is, an outcome is observed for each student in each of the two sampled subjects along with data on the teacher of the student in the given subject. Two completed teacher surveys are available for 21,324 of the 8th grade students because of some non-response and because some students did not have a class in one or both of their assigned academic subjects. The final data set consists of 42,648 observations since the unit of observation is each teacher-student pairing (Table 2).

The students participating in NELS:88 completed multiple-choice achievement tests in the subjects taught by these teachers.[8] For purposes of this analysis, the formula scores on these tests have been standardized by subject so that the changes in these scores (STEST) can be understood as effect sizes. The other outcome variables used in this study reflect the teacher's perceptions of the sampled student and the student's perception of the subject taught by that teacher (Table 2).

More specifically, this analysis focuses on three pejorative teacher assessments: whether the student was seen as frequently disruptive (DISRUPT), consistently inattentive (INATTEN), or rarely completed homework (NOHWK). The response options to these questions were simply yes or no so these three variables are binary. One potential complication with these variables is that it is not entirely clear that these outcomes should be understood as negative ones. For example, a student may become disruptive or inattentive simply because they have mastered the classroom material relative to their peers. However, the data do not support that hypothesis.

---

[8] For details on the cognitive tests, see U.S. Department of Education (1991). Test scores are, for several reasons, unavailable for roughly 4 percent of the 24,599 students who completed questionnaires. For example, some students were absent on the survey day and were only administered the questionnaire during a make-up session. Several participating schools also refused the test component of the study and test sections were not scored if a student answered fewer than 5 questions.

More specifically, using these NELS:88 data, I found that, conditional on student and subject fixed effects, students performed significantly lower on subject tests when the teacher for that subject viewed them negatively. The students viewed negatively by teachers were also substantially less likely than other students in their school to take any Advanced Placement courses over the subsequent two years and more likely to have dropped out of high school.

The remaining outcome variables used in this study are three variables reflecting the students' perception of the subject taught by the responding teacher. More specifically, students were asked whether they are afraid to ask questions in that subject, whether they look forward to their class in that subject and whether they see that subject as useful for their future. The students were given four options in response to these questions (strongly agree, agree, disagree, strongly disagree), which are coded as integers from 1 to 4. However, for ease of interpretation, the order of the responses to the "afraid" question was reversed. This implies that, for each of the three questions, higher values of the ordinal response imply a negative view of the subject. Furthermore, within each subject, the responses to each of these three questions were standardized to create the variables used in this analysis (i.e., AFASK, NOTLF, NOTUSE). Interestingly, regressions that condition on student and subject fixed effects indicate students with higher values of NOTLF, NOTUSE and AFASK have significantly lower test scores.

Table 3 presents the average values of these seven outcome measures by gender and subject. Like the NAEP data presented in Table 1, the test scores from NELS:88 indicate that girls had, on average, lower scores in math and science but higher scores in reading. Girls also had lower test scores in history. Interestingly, the magnitudes of these gender gaps in math and science are similar to those of the NAEP 13 year olds. However, the gender gap in reading scores is smaller in NELS:88 than in the NAEP. The conditional means in Table 3 also indicate that

boys are substantially more likely than girls to be viewed pejoratively by their teacher (i.e., DISRUPT, INATT, NOHWK). This pattern occurs consistently regardless of the subject. However, the pattern of self-reported student engagement with particular subjects is similar to that of test scores. More specifically, girls are more likely than boys to report higher values of NOTLF, NOTUSE and AFASK in math, science and history. However, boys report higher values of these variables with respect to English.

The remaining variables used in this study are controls for teacher and classroom observables. These include dummy variables for the gender and race-ethnicity of the teacher as well as a dummy variable that identifies whether the student shares the teacher's race-ethnicity (Table 2).[9] Another dummy variable indicates whether the teacher is state-certified in the subject they are teaching.[10] Teacher experience is measured by 10 categorical dummies (Table 2). This relatively unrestrictive approach to measuring teacher experience may important given the evidence of non-linear returns to teacher experience (e.g., Hanushek et al. 2005). The final controls capture two observable traits of the teacher's class, the number of students in the class and the percentage of students in the class who are limited English proficient (LEP).

4. SPECIFICATIONS

The specifications evaluated here are straightforward variations of those used to evaluate the labor-market returns to schooling using data on monozygotic twin pairs (e.g., Ashenfelter and Kruger 1994, Ashenfelter and Rouse 1998, Rouse 1999). More specifically, the initial specification, estimated separately for girls and boys, allows the educational outcome of student i

---

[9] Prior research suggests that a same-race may influence student outcomes through phenomenon like role-model effects, stereotype threat and teacher biases (e.g., Dee 2004, 2005, Hanushek et al. 2005).
[10] There is evidence that this field-specific qualification is associated with teacher quality (e.g., Goldhaber and Brewer 2000, Dee and Cohodes 2005).

with teacher t of subject 1 (i.e., y_{1it}) to be a function of observed student traits, $\mathbf{X_i}$ and whether the teacher of this class is female (i.e., FT_{1t}):

(1) $$y_{1it} = \boldsymbol{\alpha}\mathbf{X_i} + \beta(FT_{1t}) + \lambda\mathbf{Z_{1t}} + \mu_i + \varepsilon_{1it}$$

The terms, μ_i and ε_{1it}, are, respectively, a student fixed effect and a mean-zero error term. The term, $\mathbf{Z_{1t}}$, consists of the other observed determinants of y_{1it}, which vary at the level of the classroom and teacher. These variables include fixed effects for the subject of the class and other observed attributes of the teacher and the classroom (Table 1).

Equation (1) refers to the student when observed in either math or science. I assume that a similar specification applies when the student is observed in the second subject (i.e., social studies or reading):

(2) $$y_{2it} = \boldsymbol{\alpha}\mathbf{X_i} + \beta(FT_{2t}) + \lambda\mathbf{Z_{2t}} + \mu_i + \varepsilon_{2it}$$

One of the concerns underscored with respect to the prior literature is that a student's likelihood of being assigned to a female teacher may be correlated with the unobserved student effects (i.e., μ_i) that influence educational outcomes. The existence of this sort of non-random assignment implies that β cannot be reliably identified by evaluating equations (1) or (2) in isolation. However, the matched-pairs nature of the NELS:88 data may make it possible to identify β in such circumstances. Specifically, differencing equations (1) and (2) leads to the following:

(3) $$(y_{1it} - y_{2it}) = \beta(FT_{1t} - FT_{2t}) + \lambda(\mathbf{Z_{1t}} - \mathbf{Z_{2t}}) + (\varepsilon_{1t} - \varepsilon_{2t})$$

In addition to "first difference" (FD) estimates based on equation (3), some of the results presented here are based on stacked versions of equations (1) and (2) that condition on school fixed effects instead of student fixed effects. These more conventional OLS estimates provide some continuity with the earlier literature by examining whether unobserved student traits could

impart a bias to the estimated effect of a teacher's gender. It should also be noted that all of the inferences presented in this study are based on standard errors that accommodate heteroscedasticity clustered at the school level. This approach is arguably appropriate in light of NELS:88's clustered sampling design. Furthermore, comparisons of this approach with others (e.g., conventional standard errors as well as those that allow clustering at the student or teacher level) indicate that it implies the most conservative inferences (i.e., the largest standard errors).

The specification in equation (3) assumes that the effect of a female teacher does not depend on the subject being taught. However, this assumption may be unreasonable. For example, the magnitude of a role-model effect (or stereotype threat) could quite conceivably depend on how a teacher's gender accords with stereotypes about the particular subject matter being taught. Some of the results presented here examine this sort of heterogeneity by allowing the effect of a teacher's gender to interact with the fixed effects for the subject. Specifically, equation (1) is assumed to take the following form:

(4) $\qquad y_{1it} = \boldsymbol{\alpha X_i} + \beta_M(FTM_{1t}) + \beta_S(FTS_{1t}) + \lambda \boldsymbol{Z_{1t}} + \mu_i + \varepsilon_{1it}$

where FTM_{1t} and FTS_{1t} refer to dummy variables for whether the teacher in math and science, respectively, is female. Similarly, equation (2) can be restated as:

(5) $\qquad y_{2it} = \boldsymbol{\alpha X_i} + \beta_E(FTE_{2t}) + \beta_H(FTH_{2t}) + \lambda \boldsymbol{Z_{2t}} + \mu_i + \varepsilon_{2it}$

where FTE_{2t} and FTH_{2t} refer to dummy variables for a female teacher in English and social studies, respectively. First differencing these equations to remove the student fixed effect yields the following:

(6) $(y_{1it} - y_{2it}) = \beta_M(FTM_{1t}) + \beta_S(FTS_{2t}) + \beta_E(-FTE_{3t}) + \beta_H(-FTH_{4t}) + \lambda(\boldsymbol{Z_{1t}} - \boldsymbol{Z_{2t}}) + (\varepsilon_{1it} - \varepsilon_{2it})$

The FD estimates based on equations (3) and (6) may improve upon the prior literature by controlling for unobserved student effects in an unambiguous manner. However, it is

critically important to note that there are a number of ways that the internal validity of the FD estimates could also be compromised. For example, the FD estimates condition on unobserved student traits that are constant across subjects. However, the gender of a student's assigned teacher could be related to their *subject-specific* propensity for achievement. Furthermore, the estimated effect of a female teacher could be confounded by gender-specific patterns in a teacher's assigned classroom environments (e.g., different tracks and class sizes) and in the unobserved determinants of teacher quality. These important concerns and several related specification checks are discussed below after presenting the basic test score results.

5. GENDER AND TEST SCORES

Table 4 presents, separately for boys and girls, the estimated effect on test scores of assignment to a female teacher. The OLS specifications (i.e., columns 1 and 5) control for school and subject fixed effects as well as the teacher and classroom observables (Table 2). The first FD specification conditions only on student and subject fixed effects. The subsequent FD specifications introduce the teacher and classroom observables.

The estimates in Table 4 indicate that a female teacher has a positive but small and statistically insignificant effect on the test scores of girls. However, the FD estimates also indicate that assignment to a female teacher *reduces* the test scores of boys by a statistically significant amount of nearly .05 standard deviations. The size of this estimated effect falls somewhat in models that introduce the teacher and classroom controls. However, it should be noted that the size of these reductions is quite small (i.e., a fraction of a standard error) and that F-tests indicate that these additional regressors are not jointly significant determinants of test scores.

The comparative results from the OLS and FD specifications provide a useful indication as to whether a student's unobserved propensity for achievement is at all related to the gender of the teacher to whom they are assigned. The results for girls suggest that this is not the case; the results from the OLS and FD specifications are quite similar. However, in the case of boys' test scores, the estimated effect of a female teacher falls considerably as one moves from a model that conditions on school fixed effects to ones that control for student fixed effects. The direction of the implied bias suggests that boys with an unobserved propensity for low achievement are more likely to be assigned to male teachers. Such a non-random sorting of students could explain why prior studies have not found that a teacher's gender influences the achievement of boys (e.g., Nixon and Robinson 1999, Ehrenberg, Goldhaber and Brewer 1995).

Table 5 presents FD estimates of the effect of a female teacher in specifications that allow this effect to vary by the subject being taught (i.e., equation (6)). The results of this approach indicate that the results in Table 4 did mask some striking heterogeneity by subject, especially for girls. More specifically, these results suggest that assignment to a female teacher had statistically insignificant effects on girls' achievement in science and English. However, these estimates also suggest that assignment to a female math teacher significantly lowered girls' achievement (i.e., .061 standard deviations) while assignment to a female history teacher raised it (i.e., .074 standard deviations). These results are similar in models that introduce the controls for teacher and classroom observables. Furthermore, these differences are statistically significant. F-tests indicate that the hypothesis that a female teacher has similar effects across these four subjects can be rejected.

The results in Table 5 also suggest that the negative effects for boys' achievement of assignment to a female teacher differ considerably by subject. More specifically, these estimated

effects are particularly pronounced in math and science. However, F-tests indicate that the hypothesis that these effects are actually the same across subjects cannot be rejected at conventional levels of significance.

6. SPECIFICATION CHECKS

Taken at face value, the test-score results in Tables 4 and 5 suggest that the achievement of boys is harmed by assignment to a female teacher while the implications for girls are mixed and subject-specific. However, despite the presence of student fixed effects, these results could be quite misleading for a number of reasons. Suppose, for example, that female teachers in a particular subject are more likely to be assigned to students with a propensity for higher or lower achievement (e.g., students in a different academic track). Under such a scenario, these results could falsely suggest the existence (or absence) of educationally relevant gender dynamics between students and teachers. Similarly, these results could be biased by the presence of unobserved teacher and classroom traits (e.g., teacher quality and class size) that are associated with a teacher's gender.

It is not possible to address these concerns definitively with observational data such as these. However, there are a number of indirect ways to examine the empirical relevance of these concerns. For example, the comparative effects of a female teacher on the test scores of boys and girls can provide one way to assess the existence of bias. More specifically, the results in Table 5 indicate that a female history teacher increases girls' achievement by a statistically significant 0.074 standard deviations. The fact that female history teachers are not similarly effective in raising boys' achievement suggests that this result is not a specious reflection of unobserved teacher and classroom traits. Similarly, the fact that female science teachers appear to lower the

achievement of boys, but not girls, suggests that this estimate is not biased. And these results are similar when the few students attending single-sex schools are excluded from the sample.[11]

However, the evidence that female math teachers lower the achievement of boys and girls is more difficult to interpret. It could be that both boys and girls actually do respond negatively to female teachers of mathematics. A second candidate explanation is that these results reflect the bias imparted by unobserved teacher and classroom traits that vary by teacher's gender. For example, it could be that female math teachers are assigned to classes with fewer resources or are simply less qualified than their male counterparts.[12] A third alternative explanation is that students with a propensity for lower achievement in mathematics (e.g., those in lower tracks) are more likely to be assigned to a female math teacher.

6.1 Effects on "other subject" test scores

A straightforward counterfactual exercise provides an interesting, though ad-hoc, way to discriminate the third explanation from the first two. This test involves estimating the effect of a female *math* teacher on *science* scores. More specifically, this test involves replicating the evaluations reported in Table 5 after replacing each student's test score in math with their score in science. The premise for this test is the assumption that assignment to a female math teacher should have relatively small (or non-existent) spillover effects on science achievement if the results in Table 5 are due to gender dynamics in the classroom or the unobserved quality of the teacher or classroom. However, if assignment to a female math teacher is associated with large gains in science achievement, it would suggest that the estimates in Table 5 are biased by a

---

[11] But these cross-gender comparisons could be misleading there is a gender-specific and non-random sorting of students with a propensity for achievement in a particular subject (e.g., girls likely to excel in history being more likely to be assigned to female teacher). Some of the additional checks discussed here address this concern.

[12] Interestingly, the available data on observed teacher and classroom traits suggests that this is not the case. Specifically, auxiliary regressions indicate that female math and history teachers are not assigned to larger classes and are equally likely to have subject-specific qualifications (i.e., state certification or a subject-specific undergraduate or graduate degree).

specious correlation between the student's propensity for achievement in math and science and the likelihood of being assigned to a female math teacher.

The key results of this counterfactual exercise are presented in the first columns of Table 6. Interestingly, these results indicate that assignment to a female math teacher lowers girls' achievement in science by a weakly significant .043 standard deviations. This effect size is roughly two-thirds of the estimated direct effect of a female math teacher on mathematics achievement as reported in Table 5. The striking and somewhat implausible similarity of these effect sizes suggests that the inferences in Table 5 are, at least for girls, biased by the fact that students with a propensity for lower achievement in math and science are more likely to be assigned to female math teachers.[13]

The remaining results reported in Table 6 generalize this test to the other three subjects. More specifically, these results reflect separate specifications where student scores in science, reading and history were replaced with their respective scores in math, history and reading, while keeping the other test scores keeping the other test scores as conventionally defined (i.e., as in Table 5). In general, these results suggest that the remaining inferences in Table 5 are not biased.[14]

For example, the results from Table 5 suggest that a female history teacher significantly raised the achievement of girls. If assignment to a female history teacher had a similarly large effect on reading achievement, the validity of that inference would clearly be in doubt. However, the results in Table 6 indicate that assignment to a female history teacher has no apparent effect

---

[13] This check is, of course, not wholly dispositive because there could be genuinely large spillover effects of math achievement on science achievement. Other direct evidence on whether students with a propensity for low math achievement are more likely to be assigned to female math teachers is presented below.

[14] The one modest exception is that the estimated effect of a female science teacher on boys' math scores is similar in size to the direct effect reported in Table 5. However, this point estimate has a larger standard error and is no longer weakly significant.

on girls' reading achievement. These comparative results are consistent with the view that the apparent educational benefits received by girls assigned to female history teachers reflect something unique to those teachers or their classrooms (e.g., role model effects). As noted above, this benefit could also reflect the possibility that female history teachers are, on average, of higher quality or have more resources. However, in light of the fact that boys do not appear to benefit from female history teachers, the relevance of some gender dynamics is instead strongly suggested.

6.2 Other evidence of non-random assignment

The comparative evidence from Tables 5 and 6 suggests that this study's inferences may be compromised by the assignment to female math teachers of students (at least girls) with a propensity for lower achievement in math. Another way to examine this issue directly is to exploit a question teachers were asked about the relative achievement of a sampled students' classroom. Specifically, the teacher survey included the following question: "Which of the following best describes the achievement level of the 8th graders in this class compared with the average 8th grade student in this school?" The four options for this question included "higher levels," "average levels," "lower levels," and "widely differing."

Table 7 presents the results of auxiliary regressions, which are specified like those in Tables 4, 5 and 6 but where the dependent variable is a dummy variable for whether the teacher characterized the student's assigned class as achieving at "higher levels". The results of this exercise suggest that, in general, female teachers are less likely to be assigned to high-achieving classes but that this effect is small and statistically insignificant. However, these effects do appear to vary in a meaningful (though not statistically significant) manner by subject. More specifically, female math teachers are roughly 5 to 6 percentage points less likely to say their

class achieves at a higher level relative to others at the school. These point estimates imply a statistically significant 17 to 25 percent reduction in the mean probability.

Interestingly, these FD estimates suggest that, in the other three subjects, the teacher's gender is unrelated to the relative achievement level of the class. Overall, these results are consistent with the concern that the assignment of female math teachers to lower-achieving classes compromises this study's identification strategy. Therefore, for this remainder of this study, this analysis excludes the half of the respondents who were drawn from schools that were randomly assigned to include surveys of mathematics teachers. In other words, the resulting analysis relies only on the students and teachers from schools where the teacher surveys were sent to science and English teachers or science and social studies teachers. Because schools were randomly assigned across subject groupings, this selection should not bias the resulting inferences. However, it will reduce their statistical power.

Table 8 presents the basic test score results based on this truncated sample. The results indicate that assignment to a female teacher raises the achievement of girls by a statistically significant .042 standard deviations but lowers the achievement of boys by a similar and statistically significant amount (i.e., .046 standard deviations). For girls, there is weakly significant evidence that the test-score benefits of a female teacher are concentrated in history. In contrast, for boys, the achievement consequences of assignment to a female teacher are more uniform across subjects.

6.3 Teacher fixed effects

The previous specification checks suggest that the test score results in Table 8 reflect the gender interactions between students and teachers (e.g., role model effects or teacher biases) and not some alternative explanation. For example, these results are unlikely to reflect systematic

differences in the quality of female teachers or their classrooms because assignment to a female teacher appears to promote the achievement of female students while simultaneously harming that of male students. Similarly, the fact that assignment to a female history teacher raises girls' achievement in history but has a small and statistically significant effect on their reading achievement (Table 6) suggests that these results do not reflect the non-random assignment of girls with a propensity for achievement in those subjects. And the results from Table 7 indicate that, apart from math classes, a teacher's gender is unrelated to the perceived achievement level of their assigned students.

However, the fact that assignment to an *opposite-gender* teacher appears to have very similar achievement effects for both girls and boys suggests another, particularly compelling specification check. Specifically, in models that pool the data on boys and girls, it is possible to identify the effect of an opposite-gender (i.e., OTHSEX) teacher conditional on unrestrictive teacher fixed effects. What makes this approach practical is that, for each teacher who was surveyed by NELS:88, there were often multiple sampled students (i.e., some who did and did not share the teacher's gender). More formally, we can implement this approach by allowing the equation for science achievement to take the following form:

(7) $\qquad\qquad y_{1it} = \boldsymbol{\alpha X_i} + \beta(\text{OTHSEX}_{1it}) + \lambda \boldsymbol{Z_{1t}} + \theta_{1t} + \mu_i + \varepsilon_{1it}$

where θ_{1t} is a teacher fixed effect. Similarly, the equation for achievement in reading and social studies becomes:

(8) $\qquad\qquad y_{2it} = \boldsymbol{\alpha X_i} + \beta(\text{OTHSEX}_{1it}) + \lambda \boldsymbol{Z_{2t}} + \theta_{2t} + \mu_i + \varepsilon_{2it}.$

And first differencing these equations yields the following:

(9) $\qquad (y_{1it} - y_{2it}) = \beta(\text{OTHSEX}_{1it} - \text{OTHSEX}_{2it}) + \lambda(\boldsymbol{Z_{1t}} - \boldsymbol{Z_{2t}}) + (\theta_{1t} - \theta_{2t}) + (\varepsilon_{1it} - \varepsilon_{2it})$

The term, $\mathbf{Z_{1t}}$, consists of subject fixed effects that are specific to the gender of the student, the dummy variable for an opposite-race teacher (i.e., OTHRACE) and the two classroom variables (i.e., CLSSIZE and PCTLEP).

The key results from estimating equation (9) are reported in Table 9. The results in the first column of Table 9 are based on a version of equation (9) that excludes the teacher fixed effects. Like the results in Table 8, these baseline results indicate that assignment to an OTHSEX teacher lowers achievement by a statistically significant amount of nearly .05 standard deviations. The results in the next column reflect the introduction of fixed effects for the nearly 2300 unique teachers associated with the 8814 first-differenced observations. Not surprisingly, the introduction of these controls increases the R^2 by a factor of 6. However, the estimated effect of an OTHSEX teacher is largely unchanged (i.e., -.042) and it remains statistically significant at the 1-percent level. This result indicates that the apparent test-score consequences associated with the gender interactions between students and teachers cannot be explained by omitted teacher characteristics.[15]

7. TEACHER AND STUDENT PERCEPTIONS

The prior results suggest that assignment to an opposite-gender teacher influences the achievement of both boys and girls and that the educational relevance of these gender interactions cannot be easily explained by the unobserved characteristics of students, teachers, or classrooms. However, as noted earlier, test scores may provide a relatively narrow measure of the relevant educational outcomes associated with the gender dynamics between teachers and

---

[15] Interestingly, it is also possible to introduce fixed effects for each of the 5500 unique classrooms. And doing so leads to a similar point estimate (i.e., -.035). However, unsurprisingly, this approach also exhausts most of the sample variation associated with the 8814 first-differenced observations and leads to a dramatically increased standard error (i.e., .029).

students. In particular, the relevance of these interactions for students' intellectual engagement with particular subjects could be particularly important for understanding the subsequent patterns of course taking and achievement among older students. And teacher perceptions of student performance provide a useful complement to student achievement as measured by a low-stakes test.

The remaining inferences in this study address these issues by presenting FD estimates of how assignment to a female teacher influences teacher perceptions of a student's performance and a student's perceptions of the subject taught by a teacher (Table 2). These inferences are based on the subset of schools where math teachers were not interviewed. The results of these assessments are generally consistent with those based on test scores. For example, the FD estimates for girls are presented in Table 10. These results indicate that a given student is significantly less likely to be seen as disruptive or inattentive when with a female teacher. Furthermore, when taught by a female, girls were less likely to report that they did not look forward to a subject, that it was not useful for their future and that they were afraid to ask questions. However, these estimated effects are statistically indistinguishable from zero.

But these results also mask some interesting heterogeneity by subject. In particular, assignment to a female science teacher leads to a particularly large reduction in the probability that the student will be seen as inattentive. And, when assigned to a female science teacher, girls are significantly less likely to claim that science is not useful for their future. Combined with the results in Table 8, these estimates suggest that female science teachers are particularly effective in promoting girls' involvement with science and do so by promoting their intellectual engagement rather than by directly raising their achievement.

Table 11 presents the key results of similarly specified models when applied to the data on boys. Most of these estimates are statistically imprecise. However, these results do indicate that a given boy is significantly more likely to be seen as disruptive when assigned to a female teacher.[16] Furthermore, a boy is significantly more likely to report that he does not look forward to a particular academic subject when that subject is being taught by a female. The models that allow for interactions between a teacher's gender and the academic subject indicate that these effects are particularly pronounced in history classes.

8. CONCLUSIONS

This study's results indicate that the gender interactions between teachers and students have statistically significant effects on a diverse set of educational outcomes: test scores, teacher perceptions of student performance and student engagement with academic subjects.

And the sizes of these estimated effects are quite large. For example, assignment to an opposite-gender teacher lowers student achievement by at least .042 standard deviations. This effect size implies that just one year with a male English teacher would eliminate nearly a third of the gender gap in reading performance among 13 year olds (Table 1) and would do so by improving the performance of boys and simultaneously harming that of girls. Similarly, a year with a female teacher would close the gender gap in science achievement among 13 year olds by half and eliminate entirely the smaller achievement gap in mathematics.

The effect sizes associated with teacher and student perceptions are similarly large. For example, boys are approximately 11 percentage points more likely than girls to be seen as

---

[16] Because assignment to an OTHSEX teacher appears to have relatively similar effects on DISRUPT for both boys and girls, it is possible to evaluate the robustness of these results to conditioning on teacher fixed effects (e.g., as in Table 9). The estimate from that approach is quite similar to those reported here, indicating that these results cannot be attributed to gender patterns in teacher unobservables.

disruptive (Table 3). But the estimates presented here indicate that a year with a male teacher would close this gap by half (Tables 10 and 11). Furthermore, girls are .20 standard deviations more likely than boys to claim that science is not useful for their future (Table 3). However, the estimates presented here imply that a year with a female science teacher would close this gap by at least half (Table 10).

The sizes of these estimates suggest that the gender dynamics between teachers and middle-school students have a substantial influence on several important educational outcomes. However, how these interactions may contribute to the gender gaps in educational outcomes also depends critically on the gender distribution of teachers by subject. To examine this issue, Table 12 presents data on the percentage of teachers who are female by subject and grade. These conditional means are based on data from the more than 45,000 teachers interviewed by the U.S. Department of Education's 1999-2000 Schools and Staffing Survey.

In the 6th grade, the percentage of teachers who were female ranged from 68 to 91 percent across four subjects (i.e., math, science, reading and history). These percentages decline in later grades and for all four subjects. However, by 8th grade, most teachers were still female. For example, 83 percent of 8th grade reading teachers were female. And, interestingly, more than half of the 8th grade teachers of math and science were female as well.[17]

Combined with the test score results, these data suggest that a large fraction of boys' dramatic underperformance in reading reflects the classroom dynamics associated with the fact that their reading teachers are overwhelmingly female. For example, these estimates imply that, if half of the reading teachers in 6th, 7th, and 8th grades were male (and their effects were additive), the reading gap would fall by approximately a third. Similarly, these results also

---

[17] See Table 2.1 in U.S. Department of Education (1997) for similar data from the 1993-94 Schools and Staffing Survey.

suggest that part of boys' relative propensity to be seen as disruptive in these grades is due to the gender interactions implied by the preponderance of female teachers. However, the fact that most middle-school teachers of math, science and, to a lesser extent, history are female also implies that the gender gaps in these subjects are smaller than they would otherwise be.

The main policy implication of these results is simply to underscore that the gender interactions between students and teachers do appear to constitute an important "environmental" influence of educational outcomes for both girls and boys. However, establishing that these reduced-form effects exist does not provide explicit guidance as to the appropriate policy responses. A targeted policy response to the existence of these effects would turn critically on understanding their structural nature. For example, Sax (2005) argues for single-sex classrooms taught by teachers with gender-specific training by appealing to evidence on the differing learning styles of boys and girls. Other policy alternatives might be rooted in the conjectured existence of stereotype threat among students or of biases in the teacher behavior and expectations. What these results do suggest is that discriminating among these explanations and designing appropriately targeted policies is a promising avenue for shaping the gender patterns of educational outcomes.

REFERENCES

Abcarian, Robin. "Bright Star indeed; More than just the name of her plane, it's the new Laura Bush in her quest for a legacy." Los Angeles Times, May 7, 2005, Part E, page 1.

American Association of University Women. How Schools Shortchange Girls, Washington, DC: AAUW Educational Foundation, 1992.

Ashenfelter, Orley and Alan Krueger. "Estimating the returns to schooling using a new sample of twins," American Economic Review 84, 1994, pages 1157-1173.

Ashenfelter, Orley and Cecilia Rouse. "Income, schooling and ability: Evidence from a new sample of identifical twins," Quarterly Journal of Economics 113, 1998, pages 253-284.

Ayres, L. Laggards in our schools. New York: Russell Sage Foundation, 1909.

Bettinger, Eric P. and Bridget Terry Long. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," American Economic Review, May 2005.

Brophy, Jere. "Interactions of Male and Female Students with Male and Female Teachers," in Gender Influences in Classroom Interaction, edited by Louise Cherry Wilkinson and Cora B. Marrett, Academic Press Inc.: Orlando, FL, 1985.

Cahill, Larry. "His Brain, Her Brain," Scientific American, May 2005, 40-47.

Canes, Brandice J. and Harvey S. Rosen. "Following in her footsteps? Women's choices of college majors and faculty gender composition," Industrial and Labor Relations Review 48, 1995, 486-504.

Cavanagh, Sean. "Educators revisit girls' loss of math, science interest," Education Week, May 4, 2005.

Coley, Richard J. "Differences in the Gender Gap: Comparisons across Racial/Ethnic Groups in Education and Work," Policy Information Center, Educational Testing Service: Princeton, NJ, 2001.

Dee, Thomas S. "Teachers, race and student achievement in a randomized experiment" The Review of Economics and Statistics, February 2004, 86(1), pp. 195-210.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" American Economic Review: forthcoming.

Dee, Thomas S. and Sarah R. Cohodes. "Out-of-field teachers and student achievement: Evidence from matched-pairs comparisons," August 2005.

Dillon, Sam. "Harvard Chief Defends His Talk on Women," The New York Times, January 18, 2005.

Ehrenberg, Ronald.G., Dan D. Goldhaber and Dominic J. Brewer. "Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988." Industrial and Labor Relations Review, April 1995, 48(3), pp. 547-561.

Freeman, Catherine E. Trends in Educational Equity of Girls and Women: 2004 (NCES 2005-016). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Goldhaber, Dan D. and Dominic J. Brewer. 2000. "Does teacher certification matter? High school certification status and student achievement," Educational Evaluation and Policy Analysis 22: 129-146.

Halpern, Diane F. Sex Differences in Cognitive Abilities, 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. "The Market for Teacher Quality," NBER Working Paper No. 11154, February 2005.

Hedges, Larry V. and Nowell, Amy. "Sex differences in mental test scores, variability, and numbers of high-scoring individuals," Science 269(5520), July 7 1995, 41-45.

Ingels, Steven J., Sameer Y. Abraham, Rosemary Karr, Bruce D. Spencer and Martin R. Frankel. National Education Longitudinal Study of 1988 Base Year: Student Component Data File User's Manual. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 1990.

Jacob, Brian A. "Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education," Economics of Education Review 21(6), December 2002, 589-598.

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

Kimura, Doreen. Sex and Cognition. Cambridge and London: The MIT Press, 1999.

Kleinfeld, Judith. "The Myth that Schools Shortchange Girls: Social Science in the Service of Deception," Women's Freedom Network, Washington DC: 1998.

Lavy, Victor. "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment." NBER Working Paper No. 10678, August 2004.

Lewin, Tamar. "Ideas and Trends: How Boys Lost out to Girl Power," The New York Times, Section 4, Page 3, December 13, 1998.

Lippa, Richard A. Gender, nature, and nurture. Mahwah, NJ: Lawrence Erlbaum, 2002.

Neisser, Ulric, Gwyneth Boodoo, Thomas J. Bouchard, A. Wade Boykin, Nathan Brody, Stephen J. Ceci, Diane F. Halpern, John C. Loehlin, Robert Perloff, Robert J. Sternberg, and Susana Urbina. "Intelligence: Knowns and Unknowns," American Psychologist, February 1996, 77-101.

Neumark David and Rosella Gardecki. "Women helping women? Role model and mentoring effects on female PhD students in economics," Journal of Human Resources 33 (1), Winter 1998, 220-246.

Nixon, Lucia A. and Michael D. Robinson. "The educational attainment of young women: Role model effects of female high school faculty," Demography 36(2): May 1999, 185-194.

Riordan, Cornelius. "The Silent Gender Gap," Education Week, November 17, 1999.

Robst, John, Jack Keil and Dean Russo. "The effect of gender composition of faculty on student retention," Economics of Education Review 17(4), 429-439, 1998.

Rothstein, Donna S. "Do female faculty influence female students' educational and labor market attainments?" Industrial and Labor Relations Review 48(3), April 1995, pages 515-530.

Rouse, Cecilia. "Further estimates of the economic return to schooling from a new sample of twins," Economics of Education Review 18, 1999, pages 149-157.

Sadker, Myra and David Sadker. Failing at fairness: How our schools cheat girls. New York: Touchstone, 1994.

Sadker, David. "An Educator's Primer to the Gender War," Phi Delta Kappan 84(3), November 2002, page 235-244.

Saltzman, Amy. "Myth or Fact? Do teachers and schools favor boys to girls' detriment as some studies seem to show? Are single sex schools the only answer?" U.S. News and World Report, page 4, November 6, 1994.

Sax, Leonard S. "The Promise and Peril of Single-Sex Public Education," Education Week, March 2, 2005.

Sommers, Christina Hoff. The War Against Boys: How Misguided Feminism is Harming Our Young Men. New York: Simon and Schuster; 2000.

Spencer, Steven J., Claude M. Steele and Diane M. Quinn. "Stereotype Threat and Women's Math Performance," Journal of Experimental Social Psychology 35(1): 4-28, 1999.

U.S. Department of Education, National Center for Education Statistics, America's Teachers: Profile of a Profession,1993–94, NCES 97-460, by Robin R. Henke, Susan P. Choy, Xianglei Chen, Sonya Geis, Martha Naomi Alt, Stephen P. Broughman, Project Officer. Washington, D.C.: 1997.

U.S. Department of Education. "Psychometric Report for the NELS:88 Base Year Test Battery," NCES 91-468, by Donald A. Rock, Judith M. Pollak, Jeffrey Owings, Anne Hafner, Washington, D.C.: 1991.

Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: Massachusetts Institute of Technology, 2002.

Table 1 –1999 NAEP Scores by Subject, Age and Gender

Subject and age	Average score		Raw difference	Standardized difference
	Boys	Girls		
Science				
9 year olds	230.9	227.9	2.9 (1.7)	0.076
13 year olds	258.7	252.9	5.7 (1.4)	0.158
17 year olds	300.4	290.6	9.7 (2.2)	0.223
Math				
9 year olds	232.9	231.2	1.7 (1.3)	0.049
13 year olds	277.2	274.5	2.7 (1.4)	0.083
17 year olds	309.8	306.8	3.1 (1.7)	0.097
Reading				
9 year olds	208.5	214.8	-6.3 (2.2)	-0.161
13 year olds	253.5	265.2	-11.6 (1.8)	-0.305
17 year olds	281.5	294.6	-13.1 (2.1)	-0.314

Standard errors are reported in parentheses. Sources: U.S. Department of Education (2000) and the "NAEP 1999 Long-Term Summary Data Tables" (http://nces.ed.gov/nationsreportcard/tables/Ltt1999/).

Table 2 – Descriptive statistics, matched student-teacher observations,
8th grade students, NELS:88

Variable	Description	Mean	Standard Deviation	Sample Size
STEST	Test score in subject	0	1.0	41,271
DISRUPT	Student is frequently disruptive	.128	.334	41,580
INATT	Student is consistently inattentive	.205	.404	41,536
NOHWK	Student rarely completes homework	.198	.398	41,627
NOTUSE	Subject not useful for my future	0	1.0	40,733
NOTLF	Do not look forward to subject	0	1.0	40,839
AFASK	Afraid to ask questions in subject class	0	1.0	40,785
FT	Female teacher	.561	.496	42,648
OTHRACE	Teacher of opposite race/ethnicity	.298	.458	42,648
TBLACK	Black teacher	.079	.269	42,648
THISP	Hispanic teacher	.022	.147	42,648
TOTHER	Teacher of other race/ethnicity	.026	.160	42,648
SCERTIFD	Teacher certified by state in subject	.810	.392	42,265
TE1	Teacher experience missing	.006	.078	42,648
TE2	1-3 years of teacher experience	.107	.309	42,648
TE3	4-6 years of teacher experience	.095	.294	42,648
TE4	7-9 years of teacher experience	.100	.300	42,648
TE5	10-12 years of teacher experience	.113	.316	42,648
TE6	13-15 years of teacher experience	.126	.332	42,648
TE7	16-18 years of teacher experience	.137	.344	42,648
TE8	19-21 years of teacher experience	.100	.300	42,648
TE9	22-24 years of teacher experience	.074	.262	42,648
TE10	25+ years of teacher experience	.142	.349	42,648
CLSSIZE	Class size	23.7	6.47	41,871
PCTLEP	% of class with limited English proficiency	.012	.066	39,643
TSCI	Science class	.244	.430	42,648
TMATH	Mathematics class	.256	.436	42,648
THIST	History/social studies class	.240	.427	42,648
TENG	English class	.260	.438	42,648

Table 3 – Average student outcomes and teacher gender by subject and student gender, NELS:88

Variable	Math		Science		History		English	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
STEST	-.044	.043	-.089	.090	-.058	.060	.081	-.080
DISRUPT	.075	.184	.082	.192	.066	.173	.068	.183
INATT	.160	.264	.169	.252	.161	.237	.141	.260
NOHWK	.135	.242	.139	.256	.145	.249	.136	.277
NOTLF	.075	-.076	.085	-.088	.046	-.048	-.101	.102
NOTUSE	.039	-.040	.100	-.103	.110	-.115	-.112	.114
AFASK	.040	-.041	.016	-.016	.045	-.047	-.011	.012
TFEMALE	.582	.564	.496	.479	.418	.404	.776	.741

Table 4 – Estimated effect of a female teacher on test scores by student gender

Independent variable	Girls				Boys			
	OLS	FD	FD	FD	OLS	FD	FD	FD
Female teacher	.018	.007	.010	.013	-.016	-.047‡	-.037†	-.040†
	(.017)	(.013)	(.013)	(.015)	(.020)	(.014)	(.015)	(.016)
R^2	.3561	.0249	.0281	.0292	.3579	.0072	.0081	.0091
Sample size	19167	10361	10166	8999	19004	10255	10074	8885
p-value (F test)	-	-	.1655	.1682	-	-	.6043	.6226
School fixed effects?	yes	no	no	no	yes	no	no	no
Student fixed effects?	no	yes	yes	yes	no	yes	yes	yes
Teacher controls?	yes	no	yes	yes	yes	no	yes	yes
Classroom controls?	yes	no	no	yes	yes	no	no	yes

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects. The OLS models include student-level dummies for race-ethnicity and SES quartile. The p-value refers to an F-test of the joint significance of the teacher or classroom controls added to the given specification.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 5 – FD estimates of the effect of a female teacher on test scores by student gender and academic subject

Independent variable	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher in math	-.061†	-.061†	-.066†	-.082‡	-.073‡	-.078‡
	(.026)	(.026)	(.027)	(.027)	(.027)	(.028)
Female teacher in science	-.010	-.002	.017	-.063†	-.051*	-.048*
	(.026)	(.026)	(.027)	(.027)	(.027)	(.028)
Female teacher in English	.044	.040	.041	-.041	-.035	-.038
	(.028)	(.029)	(.031)	(.027)	(.028)	(.030)
Female teacher in history	.074‡	.078‡	.075‡	.004	.017	.010
	(.026)	(.026)	(.027)	(.031)	(.031)	(.033)
R^2	.0274	.0305	.0317	.0081	.0081	.0100
Sample size	10316	10166	8999	10255	10074	8885
p-value (H$_0$: $\beta_M = \beta_S = \beta_E = \beta_H$)	.0014	.0013	.0021	.1810	.1640	.2373
Teacher controls?	no	yes	yes	no	yes	yes
Classroom controls?	no	no	yes	no	no	yes

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects. The p-value refers to an F-test of the hypothesis that the four coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 6 – Counterfactual FD estimates of the effect of a female teacher on "other-subject" test scores
by student gender and academic subject

Independent variable	Math score replaced by science		Science score replaced by math		Reading score replaced by history		History score replaced by reading	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Female teacher in math	-.043*	-.030	-.070†	-.073‡	-.058†	-.078†	-.080‡	-.069†
	(.026)	(.025)	(.028)	(.028)	(.028)	(.031)	(.027)	(.027)
Female teacher in science	.020	-.053*	-.036	-.046	.005	-.026	.003	-.038
	(.027)	(.028)	(.032)	(.032)	(.027)	(.030)	(.027)	(.028)
Female teacher in English	.031	-.040	.043	-.019	.013	-.026	.042	-.036
	(.033)	(.030)	(.031)	(.031)	(.031)	(.038)	(.031)	(.030)
Female teacher in history	.081‡	.005	.045	.021	.074‡	.012	-.006	-.007
	(.026)	(.031)	(.032)	(.036)	(.027)	(.033)	(.027)	(.029)
R^2	.0315	.0070	.0267	.0118	.0094	.0073	.0084	.0042
p-value (H_0: $\beta_M = \beta_S = \beta_E = \beta_H$)	.0045	.4960	.0090	.1857	.0077	.2507	.0176	.4819
Sample size	8998	8887	9000	8886	8987	8858	9025	8903

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the teacher and classroom controls. The p-value refers to an F-test of the hypothesis that the four coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 7 –Auxiliary regressions, FD estimates of the effect of a female teacher on assignment to a high-achieving class by student gender and academic subject

Independent variable	Girls		Boys	
Female teacher	-.013 (.014)	-	-.016 (.014)	-
Female teacher in math	-	-.059† (.026)	-	-.047† (.024)
Female teacher in science	-	.024 (.027)	-	.020 (.025)
Female teacher in English	-	-.010 (.027)	-	-.018 (.028)
Female teacher in history	-	-.003 (.028)	-	-.016 (.026)
R^2	.0084	.0051	.0090	.0105
p-value (H$_0$: $\beta_M = \beta_S = \beta_E = \beta_H$)	-	.1508	-	.2741
Dependent mean	.24		.27	
Sample size	9261	9261	9191	9191

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the teacher and classroom controls. The p-value refers to an F-test of the hypothesis that the four coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 8 –FD estimates of the effect of a female teacher on test scores by student gender and academic subject, excluding mathematics

Independent variable	Girls		Boys	
Female teacher	.042† (.020)	-	-.046† (.022)	-
Female teacher in math	-	-	-	-
Female teacher in science	-	.025 (.027)	-	-.054* (.029)
Female teacher in English	-	.026 (.046)	-	-.020 (.045)
Female teacher in history	-	.090† (.037)	-	-.051 (.041)
R^2	.0400	.0408	.0067	.0069
p-value (H$_0$: $\beta_S = \beta_E = \beta_H$)	-	.0824	-	.1905
Sample size	4426	4426	4322	4322

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the teacher and classroom controls. The p-value refers to an F-test of the hypothesis that the three coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 9 – FD estimates of the effect of an opposite-gender teacher on test scores, pooled data on boys and girls, excluding mathematics

Independent variable	(1)	(2)
Opposite-gender teacher	-.048‡ (.014)	-.042‡ (.016)
R^2	.0410	.2621
Sample size	8814	8814
Teacher fixed effects?	no	yes

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects specific to the student's gender, a dummy variable for students observed with a teacher of a different race/ethnicity and the classroom controls. Model (1) also controls for the gender and race-ethnicity of the teacher.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 10 – FD estimates of the effect of a female teacher on girls' non-test outcomes, excluding mathematics

| Dependent variable | Female teacher | By academic subject | | | p-value |
		Science	English	History	(H_0: $\beta_S = \beta_E = \beta_H$)
DISRUPT	-.021† (.010)	-.027† (.012)	-.004 (.026)	-.021 (.016)	.1088
INATT	-.032‡ (.013)	-.042† (.018)	-.006 (.032)	-.030 (.024)	.0597
NOHWK	.025† (.010)	.028* (.015)	.025 (.024)	.018 (.021)	.1277
NOTLF	-.068 (.042)	-.113† (.055)	-.006 (.099)	-.020 (.080)	.2104
NOTUSE	-.036 (.030)	-.107† (.042)	.075 (.071)	.027 (.051)	.0439
AFASK	-.007 (.030)	.009 (.040)	-.017 (.060)	-.033 (.056)	.9228

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the controls for teacher and classroom observables. The p-value refers to an F-test of the hypothesis that the three coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 11 – FD estimates of the effect of a female teacher on boys' non-test outcomes, excluding mathematics

Dependent variable	Female Teacher	By academic subject			p-value $(H_0: \beta_S = \beta_E = \beta_H)$
		Science	English	History	
DISRUPT	.033†	.012	.058†	.058†	.0188
	(.013)	(.017)	(.029)	(.024)	
INATT	.002	-.020	.036	.021	.3943
	(.015)	(.021)	(.033)	(.027)	
NOHWK	.030*	.019	.016	.064†	.1029
	(.016)	(.020)	(.035)	(.027)	
NOTLF	.093†	.018	.099	.247‡	.0025
	(.043)	(.051)	(.110)	(.067)	
NOTUSE	.009	-.050	.124	.042	.2069
	(.034)	(.042)	(.085)	(.058)	
AFASK	.042	.029	.031	.078	.4300
	(.030)	(.037)	(.053)	(.051)	

Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects and the controls for teacher and classroom observables. The p-value refers to an F-test of the hypothesis that the three coefficients are equal.
* Statistically significant at the 10-percent level
† Statistically significant at the 5-percent level
‡ Statistically significant at the 1-percent level

Table 12 – Percent of teachers who are female by grade and subject,
1999-2000 Schools and Staffing Survey (SASS)

| Grade | Subject | | | |
	Science	Reading	Math	History
6	0.681	0.908	0.717	0.575
7	0.627	0.871	0.647	0.495
8	0.581	0.831	0.648	0.477
9	0.461	0.742	0.520	0.319
10	0.469	0.732	0.523	0.330
11	0.448	0.717	0.524	0.332
12	0.441	0.711	0.524	0.324

Notes: These conditional means are based on the author's calculations using SASS respondents who are regular full and part-time teachers (n = 46,542). The subject designations are based on each teacher's main teaching assignment. Sample weights are used.