Teachers and the Gender Gaps in Student Achievement

Thomas S. Dee

ABSTRACT

A prominent class of explanations for the gender gaps in student outcomes focuses on the interactions between students and teachers. In this study, I examine 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 study, which provides contemporaneous data on student outcomes in two different subjects. Within-student comparisons 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.

I. 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 the heated discussion, both in popular and academic settings, has focused on assessing the relative contributions of biological and environmental determinants (that is, nature versus nurture). 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

Thomas S. Dee is an associate professor of economics and Director of the Program in Public Policy at Swarthmore College and a Faculty Research Fellow, NBER. *This research was supported by the Spencer Foundation and by a grant from the American Educational Research Association which receives funds for its AERA Grants Program from the National Science Foundation and the National Center for Education Statistics of the Institute of Education Sciences (U.S. Department of Education) under NSF Grant #REC-0310268. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies. The author would also like to thank the referees and seminar participants at Stanford University, the College of William and Mary and Drexel University for helpful comments. The data used in this article can be obtained beginning January 2008 through December 2010 from Thomas S. Dee, 500 College Avenue, Department of Economics, Swarthmore College, Swarthmore, PA 19081 dee@swarthmore.edu.*

[Submitted March 2006; accepted September 2006]

ISSN 022-166X E-ISSN 1548-8004

contrasting images of how interactions with teachers may shape the relative cognitive development and intellectual engagement of boys and girls (AAUW 1992; Sadker and Sadker 1994; Saltzman 1994; Kleinfeld 1998; Lewin 1998; Sommers 2000; and Sadker 2002). This study presents new evidence on the educational consequences of student-teacher interactions by assessing the effects of a same-gender teacher on a variety of student outcomes.

Recent discussions of gender interactions within classrooms have centered on controversial claims that teachers consistently privilege boys over girls (for example, with more positive feedback and helpful questions). However, the literature on student-teacher interactions has also focused whether student outcomes 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 also may communicate different (and self-fulfilling) expectations to the boys and girls in their classrooms (that is, Pygmalion effects). Prior studies have examined the empirical relevance of such gender interactions by assessing the reduced-form impact of assignment to a same-gender teacher on educational outcomes (for example, Bettinger and Long 2005; Canes and Rosen 1995; Ehrenberg, 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 (that is, eighth graders) instead of students at secondary or postsecondary levels. The distinction between younger and older students may be particularly relevant because early adolescence corresponds more closely with the age at which the gender gaps in educational achievement become particularly pronounced (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 also may 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 (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 eighth graders. However, NELS:88 also surveyed *two* of each student's academic-subject teachers. These surveys elicited information

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

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

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

on each teacher's background and on how each teacher viewed the performance of the sampled student. Furthermore, NELS:88 also gathered self-reported information from students on their attitudes toward the teacher's academic subject and fielded student tests in each academic subject. The cluster-sample (Wooldridge 2002, page 328) nature of the NELS:88 data 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 (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. These matched-pairs comparisons control for subject fixed effects and student fixed effects that are constant across subjects. However, the internal validity of such within-student comparisons could still be compromised by the nonrandom sorting by students with subject-specific propensities for achievement and by unobserved teacher and classroom traits correlated with gender. Evidence on the empirical relevance of these potential sources of specification error is presented.

II. 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 3^{rd} grade, boys have slightly higher mathematics scores and lower reading scores. The 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 nine-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 nine year-olds, it is the only difference that is statistically significant.

However, the NAEP data indicate that, over the next four years, each of the subject-specific achievement gaps increases. Specifically, as we move from nine-year-olds 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.

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 (that is, 0.3 standard deviations) is nearly half the size of the corresponding black-white differential, a topic that 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 (that is, 0.2 and 0.1 standard deviations, respectively) is more modest but still qualitatively large (for example, 20 and 10 percent of the respective black-white gaps).

The gender differences in 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 the variance in female test scores, 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 (Jacob 2002). 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). However, female college students continue to be underrepresented in certain technical fields like engineering and 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 Bell Curve'' suggested that biological factors do contribute to the gender differences in skills (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 (Kimura 1999; Halpern 2000; Lippa 2002; and Cahill 2005). However, Neisser et al. (1996) note that these biological differences interact with environmental factors that appear soon after birth. Furthermore, discussions of the evidence on gender differences frequently emphasize that there is ''substantial overlap in the distribution of male and female scores'' (Coley 2001).

The gender dynamics in classrooms are also frequently portrayed as an important environmental source of the gender differences in educational outcomes (for example, 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 the 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 (AAUW 1992, Sadker and Sadker 1994, Saltzman 1994, Kleinfeld 1998, Lewin 1998, and Sommers 2000). Similarly, cognitive process theories (Jones and Dindia 2004) suggest that teachers may subtly communicate that they have different academic expectations of boys and girls. The biased expectations of teachers may then become self-fulfilling when students respond to them (that is, a Pygmalion effect).

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 that examine teacher biases and conclude that a teacher's gender is ''the most obvious factor that seems to shape sex equity in the classroom.'' However, those 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 the 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. The recent literature on the phenomenon known as stereotype threat provides another perspective on how students might react to a teacher's gender. Stereotype threat 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 the structural explanations of student-teacher interactions but instead provides reduced-form evidence on the educational consequences of assignment to a same-gender teacher. Interestingly, most prior evidence on the effects of a same-gender teacher has focused on postsecondary and graduate settings (for example, 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 tenth 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 nonblind 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 the literature on the gender interactions between students and teachers in three ways. First, this study focuses on younger students (that is, eighth 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 (that is, 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 the context of student-teacher interactions is that they may be biased by the nonrandom 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 downward. 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. A fixed-effects approach is feasible because of a unique, matched-pairs feature of a major longitudinal study, which is described in the next section.

III. 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 eighth 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 eighth grade enrollment. Approximately 26 students were then randomly chosen and surveyed within each participating school. However, approximately five percent of students did not participate in the student survey because they were unable to understand or complete the survey materials. Spencer et al. (1990, Section 2.1.1) find that virtually all of the exclusions were due to mental disabilities and language barriers.

NELS:88 also fielded questionnaires to the teachers responsible for teaching each of the selected students in two of four academic subjects: mathematics, science, English, and history. The two surveyed teachers were chosen by randomly assigning each sampled school to one of four subject-area groupings: mathematics/English, mathematics/history, science/English, and science/history. 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 administered tests in each of the four academic subjects and surveyed students about their perception of each subject.

The student and teacher surveys in NELS:88 imply that we observe a variety of student-level educational outcomes (that is, test scores in each academic subject, teacher perceptions of individual students, and student perceptions of specific academic subjects). 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 eighth grade students because of some nonresponse 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 because the unit of observation is each teacher-student pairing (Table 2).

The students participating in NELS:88 completed multiple-choice achievement tests in mathematics, science, reading and history.[1] For purposes of this analysis, the formula scores on the 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).

1. For details on the cognitive tests, see Rock et al. (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 makeup 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. Fortunately, auxiliary regressions indicate that the absence of a subject test is unrelated to the gender of the student's teacher in that subject.

Table 2
Descriptive Statistics, Matched Student-Teacher Observations, Eighth 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	0.128	0.334	41,580
INATT	Student is consistently inattentive	0.205	0.404	41,536
NOHWK	Student rarely completes homework	0.198	0.398	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	0.561	0.496	42,648
OTHRACE	Teacher of opposite race/ethnicity	0.300	0.458	42,648
TBLACK	Black teacher	0.079	0.269	42,648
THISP	Hispanic teacher	0.022	0.147	42,648
TOTHER	Teacher of other race/ethnicity	0.025	0.160	42,648
SCERTIFD	Teacher certified by state in subject	0.810	0.392	42,265
TE1	Teacher experience missing	0.006	0.078	42,648
TE2	1–3 years of teacher experience	0.107	0.309	42,648
TE3	4–6 years of teacher experience	0.095	0.294	42,648
TE4	7–9 years of teacher experience	0.100	0.300	42,648
TE5	10–12 years of teacher experience	0.113	0.316	42,648
TE6	13–15 years of teacher experience	0.126	0.332	42,648
TE7	16–18 years of teacher experience	0.137	0.344	42,648
TE8	19–21 years of teacher experience	0.100	0.300	42,648
TE9	22–24 years of teacher experience	0.074	0.262	42,648
TE10	25+ years of teacher experience	0.142	0.349	42,648
CLSSIZE	Class size	23.7	6.47	41,871
PCTLEP	Percent of classmates with limited English proficiency	0.012	0.066	39,643
TSCI	Science class	0.244	0.430	42,648
TMATH	Mathematics class	0.256	0.436	42,648
THIST	History/social studies class	0.240	0.427	42,648
TENG	English class	0.260	0.438	42,648

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 *DISRUPT*, *INATTEN* and *NOHWK* 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 the hypothesis that these teacher perceptions actually reflect higher levels of achievement. More specifically, using the NELS:88 data, I found that, conditional on student and subject fixed effects, students performed significantly lower on subject tests when the teacher for the 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 four 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 the subject, whether they look forward to their class in the subject and whether they see the 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 question about being afraid to ask questions was reversed. This coding scheme 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 (that is, *AFASK*, *NOTLF*, and *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.

The subject-specific gender gaps in the NELS:88 data are similar to those based on the NAEP data (Table 1) with boys outperforming girls in math, science, and history but underperforming in reading. However, the data on teacher perceptions suggest a different pattern of gender gaps across subjects. More specifically, boys are substantially more likely than girls to be viewed pejoratively by their teacher (that is, higher values of *DISRUPT*, *INATT*, and *NOHWK*) regardless of the subject. Interestingly, the gender differences in self-reported student engagement with particular subjects are similar to the gender differences in 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 *NOTLF*, *NOTUSE*, and *AFASK* 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). Recent research suggests that a same-race teacher may influence student outcomes through phenomenon like role-model effects, stereotype threat, and teacher biases (Dee 2004, 2005; Hanushek et al. 2005). Another dummy variable indicates whether the teacher is state-certified in the subject they are teaching. Recent evidence (for example, Goldhaber and Brewer 2000; Dee and Cohodes, forthcoming) indicates that such field-specific, state certification is associated with teacher quality. Teacher experience is measured by 10 categorical dummies (Table 2). This relatively unrestrictive approach to measuring teacher experience may be important given the evidence of nonlinear returns to teacher experience (Hanushek et al. 2005). The final controls, which are drawn from the teacher surveys, 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).

IV. Specifications

The specifications evaluated here are straightforward variations of the research designs used to evaluate the labor market returns to schooling using data on monozygotic twin pairs (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* with teacher *t* of subject 1 (that is, y_{1it}) to be a function of observed student traits, X_i and whether the teacher of the class is female (that is, FT_{1t}):

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

The terms, μ_i and ϵ_{1it}, are, respectively, a student fixed effect and a mean-zero error term. The term, 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 (that is, history or reading):

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

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 (that is, μ_i) that influence educational outcomes. The existence of such nonrandom 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:

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

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 heteroskedasticity clustered at the school level. This school-level clustering is arguably appropriate in light of NELS:88's sampling design. Furthermore, comparisons of this approach with others (for example, conventional standard errors as well as standard errors that allow clustering at the student or teacher level) indicate that it implies the most conservative inferences (that is, 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, the assumption of a common treatment effect 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 subject-specific 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:

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

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:

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

where FTE_{2t} and FTH_{2t} refer to dummy variables for a female teacher in English and history, respectively. First differencing Equations 4 and 5 to remove the student fixed effect yields the following:

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

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 also could 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 (for example, different tracks and class sizes). Similarly, it may be that unobserved teacher quality differs consistently by gender. These important concerns and several related specification checks are discussed below after presenting the basic test score results.

V. Gender and Test Scores

Table 3 presents, separately for boys and girls, the estimated effect on test scores of assignment to a female teacher. The OLS specifications (Columns 1 and 5) control for school and subject fixed effects as well as student traits (race, ethnicity and SES quartile) and 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 3 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 0.05 standard deviations. The size of the

Table 3
Estimated Effect of a Female Teacher on Test Scores by Student Gender

	Girls				Boys			
Independent variable	OLS	FD	FD	FD	OLS	FD	FD	FD
Female teacher	0.018	0.007	0.010	0.014	−0.016	−0.047***	−0.037**	−0.040**
	(0.017)	(0.013)	(0.013)	(0.015)	(0.020)	(0.014)	(0.015)	(0.016)
R^2	0.3551	0.0158	0.0189	0.0200	0.3593	0.0131	0.0138	0.0150
Sample size	19,167	10,316	10,166	8,999	19,004	10,255	10,074	8,885
p-value (F-test)	—	—	0.1976	0.1836	—	—	0.6799	0.5938
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

Note: 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 5 percent level and *** Statistically significant at the 1 percent level

estimated effect of a female teacher falls somewhat in models that introduce the teacher and classroom controls. However, it should be noted that the size of the reductions in the estimated effect of a female teacher is relatively small (that is, a fraction of a standard error). Furthermore, F-tests indicate that the additional regressors are not jointly significant determinants of test scores. It should also be noted that the results are quite similar when these regressions are uniformly limited to the observations for which the teacher and classroom observables are available.

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 the propensity to be assigned to a female teacher is unrelated to the unobserved determinants of achievement; 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 nonrandom sorting of students could explain why prior studies (for example, Nixon and Robinson 1999; Ehrenberg, Goldhaber, and Brewer 1995) have not found that a teacher's gender influences the achievement of boys.

Table 4 presents FD estimates of the effect of a female teacher in specifications that allow this effect to vary by the subject being taught (that is, the estimated coefficients on FTM_{1t}, FTS_{1t}, $-FTE_{2t}$, and $-FTH_{2t}$ in Equation 6). The results in Table 4 indicate that the results in Table 3 did mask some striking heterogeneity by subject, especially for girls. More specifically, the results in Table 4 suggest that assignment to a female teacher had statistically insignificant effects on girls' achievement in science and English. However, assignment to a female math teacher significantly lowered girls' achievement by 0.061 standard deviations while assignment to a female history teacher raised it by 0.074 standard deviations. The results for girls are similar in models that introduce the controls for teacher and classroom observables. Furthermore, F-tests indicate that the hypothesis that a female teacher has similar effects across these four subjects can be rejected.

The results in Table 4 also suggest that the negative effects for boys' achievement of assignment to a female teacher differ considerably by subject. More specifically, the estimated effects of a female teacher are particularly pronounced in math and science. However, F-tests indicate that the hypothesis that these effects are actually the same across all four subjects cannot be rejected at conventional levels of significance.

VI. Specification Checks

Taken at face value, the test-score results in Tables 3 and 4 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

Table 4

FD estimates of the Effect of a Female Teacher on Test Scores by Student Gender and Academic Subject

	Girls			Boys		
Teacher trait	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher in math	−0.061**	−0.060**	−0.065**	−0.081***	−0.072***	−0.076***
	(0.026)	(0.026)	(0.027)	(0.027)	(0.027)	(0.028)
Female teacher in science	−0.010	−0.002	0.017	−0.063**	−0.051*	−0.049*
	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)	(0.029)
Female teacher in English	0.043	0.039	0.039	−0.041	−0.035	−0.038
	(0.028)	(0.029)	(0.031)	(0.027)	(0.028)	(0.030)
Female teacher in history	0.074***	0.079***	0.077***	0.004	0.016	0.010
	(0.026)	(0.026)	(0.027)	(0.032)	(0.032)	(0.033)
R^2	0.0183	0.0214	0.0225	0.0140	0.0147	0.0158
Sample size	10,316	10,166	8,999	10,255	10,074	8,885
p-value (H_0: $\beta_M = \beta_S = \beta_E = \beta_H$)	0.0013	0.0013	0.0021	0.1942	0.1640	0.2513
Teacher controls?	no	yes	yes	no	yes	yes
Classroom controls?	no	no	yes	no	no	yes

Note: 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; and *** Statistically significant at the 1 percent level.

likely to be assigned to students with a propensity for higher or lower achievement (for example, students in a different academic track). Under such a scenario, the results in Tables 3 and 4 could falsely suggest the existence (or absence) of educationally relevant gender dynamics between students and teachers. Similarly, the results in Tables 3 and 4 could be biased by the presence of unobserved teacher and classroom traits (for example, teacher quality and class size) that are associated with a teacher's gender.

It is not possible to address these concerns definitively with the available observational data. However, there are a number of indirect ways to examine the empirical relevance of the possible sources of bias. 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 4 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 the effect of female teachers on girls' history achievement 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 the estimated effect of female science teachers on boys' achievement is not biased by unobserved teacher and classroom traits.[2] Furthermore, I found that the estimated effects of a female teacher were similar when the few sampled students attending single-sex schools were excluded from the sample.

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 the results in Table 4 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. Interestingly, the available data on observed teacher and classroom traits suggest that this is not the case. In particular, auxiliary regressions indicate that female math and history teachers are not assigned to larger classes and are equally likely to have subject-specific qualifications (that is, state certification or a subject-specific undergraduate or graduate degree in the subject they teach).[3] A third alternative explanation for the results in Table 4 is that students with a propensity for lower achievement in mathematics (for example, students in lower tracks) are more likely to be assigned to a female math teacher. The descriptive evidence on the patterns of student ability-grouping across academic subjects during middle school suggests that the third explanation is a possibility. In particular, ability grouping is common in the eighth grade mathematics curriculum, which often

2. However, cross-gender comparisons could be misleading there is a gender-specific and nonrandom sorting of students with a propensity for achievement in a particular subject (for example, 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.

3. Whether observed teacher traits actually reflect teacher quality is a controversial topic. However, recent studies (Dee and Cohodes, forthcoming, and Goldhaber and Brewer 2000) suggest that subject-specific teacher qualifications (that is, state certification or a subject-specific undergraduate or graduate degree) do promote student achievement.

directs more advanced students toward classes in pre-algebra and algebra (Loveless 1998).[4]

A. 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 4 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 nonexistent) spillover effects on science achievement if the results in Table 4 are due to gender dynamics in the math classroom or the unobserved quality of the math 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 4 are biased by a specious correlation between the student's propensity for achievement in math and science and the likelihood of being assigned to a female math teacher.

Interestingly, the results of such a regression indicate that assignment to a female math teacher lowers girls' achievement in science by 0.042 standard deviations, an estimate that falls just short of weak statistical significance (p-value $= 0.104$). This estimated effect is roughly two-thirds of the estimated direct effect of a female math teacher on mathematics achievement as reported in Table 4. The estimated effect of a female math teacher on the science achievement of boys is also relatively large (-0.030) but statistically insignificant. Nonetheless, the somewhat implausible comparative size of the effect of female math teachers on science achievement suggests that the inferences in Table 4 could (at least for girls) be 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. The evidence from the estimated effects of female math teachers on science achievement is, of course, not wholly dispositive because of the lack of statistical power and because there could be genuinely large spillover effects of math achievement on science achievement. Other more 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.

However, I also generalized this counterfactual exercise to the other subjects by evaluating 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 (that is, as in Table 4). The results of this exercise suggested that the remaining inferences in Table 4 are not biased. For example, the results from Table 4 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, this exercise indicated that assignment to a female history teacher had a quite small (-0.006) and statistically

4. Ability grouping of middle school students is also common in English but not in science or history (Loveless 1998; Hoffer 1992).

insignificant (p-value = 0.827) effect 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 (for example, role model effects). As noted above, the achievement gains of girls assigned to female history teachers also could 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 within classrooms is instead strongly suggested.

B. Other Evidence of Nonrandom Assignment

Another way to assess whether female math teachers are more likely to be assigned to students with a propensity for lower achievement in math would be to examine other data on the character of teacher assignments by gender and subject. While NELS-88 did not field explicit questions on teacher assignments, 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 eighth graders in this class compared with the average eighth grade student in this school?" The four options for this question included "higher levels," "average levels," "lower levels," and "widely differing."

On average, about 25 percent of teachers characterized the sampled students' class as high-achieving. Table 5 presents the results of some auxiliary regressions that examine whether the achievement level of an assigned class differed by academic subject

Table 5
Auxiliary Regressions, Estimated of a Female Teacher on Assignment to a High-Achieving Class by Academic Subject

Teacher trait	(1)	(2)	(3)	(4)
Female teacher	0.002	—	0.004	—
	(0.012)		(0.013)	
Female teacher in math	—	−0.043**	—	−0.038
		(0.021)		(0.024)
Female teacher in science	—	0.025	—	0.008
		(0.022)		(0.021)
Female teacher in English	—	−0.001	—	0.028
		(0.027)		(0.029)
Female teacher in history	—	0.028	—	0.028
		(0.026)		(0.023)
School fixed effects	no	no	yes	yes

Note: Standard errors, adjusted for school-level clustering, are reported in parentheses. All models include subject fixed effects. ** Statistically significant at the 5 percent level.

and teacher's gender. Specifically, the dependent variable is a dummy variable for whether the teacher characterized the assigned class as achieving at "higher levels." The results of this exercise suggest that, in general, female teachers are more likely to be assigned to high-achieving classes but that this effect is quite small and statistically insignificant. However, the estimated effects of female teachers do appear to vary in a meaningful manner by subject. With respect to reading, science, and history, the estimates in Table 5 suggest that a teacher's gender is unrelated to the relative achievement level of their assigned students. However, female math teachers were approximately four percentage points less likely to say their class achieves at a higher level relative to others at the school. This point estimate is fairly large, implying a reduction in the mean probability of about 17 percent. However, it should be noted that the estimate falls just short of weak statistical significance (p-value $=$ 0.118) in models that introduce school fixed effects.

Nonetheless, the evidence discussed here is quite suggestive of the concern that female math teachers are more likely to be assigned to lower-achieving classes. Such a propensity in teacher assignments could compromise this study's identification strategy. Therefore, for the 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 history teachers. Because schools were randomly assigned across subject groupings, an analysis based on the remaining schools should not bias the resulting inferences. However, it will reduce their statistical power.

Table 6 presents the basic test score results based on the truncated sample. The results indicate that assignment to a female teacher raises the achievement of girls by a statistically significant 0.045 standard deviations but lowers the achievement of boys by a similar and statistically significant amount (that is, 0.047 standard deviations). The results by subject suggest that, for girls, the test-score benefits of a female teacher are concentrated in history. However, the null hypothesis that the three coefficients are equal could not be rejected. The results in Table 6 suggest that, for boys, the achievement consequences of assignment to a female teacher are more uniform across subjects.

The results in Table 6 control for the key teacher and classroom observables available in the teacher surveys. However, the results in Table 6 could be biased if female teachers had predominately female students and the resulting differences in the gender composition of student peers in turn shaped student outcomes. However, the NELS:88 data indicate that the gender composition of students varies by only two percentage points across male and female teachers. The effects associated with peers' gender would have to be implausibly large for the small difference in the gender composition of classrooms to explain away the results in Table 6. Nonetheless, I used the student data in NELS:88 to construct a measure of the percent female in each sampled classroom. This is an admittedly noisy measure; however, I did find that the results in Table 6 were robust to including it as a control. I also examined whether the results in Table 6 masked any response heterogeneity by students' race, ethnicity, and socioeconomic status. In general, the results were similar across subgroups defined by such traits with the exception that the test score gains

Table 6
FD Estimates of the Effect of a Female Teacher on Test Scores by Student Gender and Academic Subject, Excluding Mathematics

Teacher trait	Girls		Boys	
Female teacher	0.045**	—	−0.047**	—
	(0.020)		(0.022)	
Female teacher in science	—	0.027	—	−0.055*
		(0.027)		(0.029)
Female teacher in English	—	0.023	—	−0.021
		(0.046)		(0.045)
Female teacher in history	—	0.097***	—	−0.052
		(0.037)		(0.041)
R^2	0.0304	0.0313	0.0105	0.0106
p-value (H_0: $\beta_S = \beta_E = \beta_H$)	—	0.2111	—	0.8073
Sample size	4,426	4,426	4,322	4,322

Note: 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; and *** Statistically significant at the 1 percent level.

associated with assignment to a female teacher were particularly large for Hispanic females.

C. Teacher Fixed Effects

The previous specification checks suggest that the test score results in Table 6 reflect the gender interactions between students and teachers (for example, role model effects or teacher biases) and not some alternative explanation. For example, the results in Table 6 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 the achievement of male students. Similarly, the fact that assignment to a female history teacher raises girls' achievement in history but has a small and statistically insignificant effect on their reading achievement suggests that the results in Table 6 do not reflect the nonrandom assignment of girls with a propensity for achievement in those subjects. And the results from Table 5 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 (that is, *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

Table 7
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	−0.048***	−0.043***
	(0.014)	(0.016)
R^2	0.0349	0.2594
Sample size	8,814	8,814
Teacher fixed effects?	no	yes

Note: 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 1 percent level.

often multiple sampled students (that is, 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:

$$(7) \quad y_{1it} = \alpha X_i + \beta(OTHSEX_{1it}) + \lambda Z_{1t} + \theta_{1t} + \mu_i + \varepsilon_{1it}$$

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

$$(8) \quad y_{2it} = \alpha X_i + \beta(OTHSEX_{2it}) + \lambda Z_{2t} + \theta_{2t} + \mu_i + \varepsilon_{2it}.$$

And first differencing Equations 7 and 8 yields the following:

$$(9) \quad (y_{1it} - y_{2it}) = \beta(OTHSEX_{1it} - OTHSEX_{2it}) + \lambda(Z_{1t} - Z_{2t}) + (\theta_{1t} - \theta_{2t}) + (\varepsilon_{1it} - \varepsilon_{2it})$$

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

The key results from estimating Equation 9 are reported in Table 7. The results in the first column of Table 7 are based on a version of Equation 9 that excludes the teacher fixed effects. Like the results in Table 6, the baseline results that do not condition on teacher fixed effects indicate that assignment to an *OTHSEX* teacher lowers achievement by a statistically significant amount of nearly 0.05 standard deviations. The results in the next column reflect the introduction of fixed effects for the more than 2,100 unique teachers associated with the 8,814 first-differenced observations. Not surprisingly, the introduction of these controls increases the R^2 substantially (more specifically, by a factor of more than seven). However, the estimated effect of an *OTHSEX* teacher remains largely unchanged (that is, -0.043) and it remains statistically significant at the 1 percent level. The comparative results in Tables 6 and 7 suggest that the apparent test-score consequences associated with the gender interactions between students and teachers cannot be explained by omitted teacher characteristics.[5]

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

Tables 8 and 9 present 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. The estimates in Tables 8 and 9 are based on the subset of schools where math teachers were not interviewed and are generally consistent with the results based on test scores (Tables 6 and 7). For example, the FD estimates for girls are presented in Table 8. The results in Table 8 indicate that female students are significantly less likely to be seen as disruptive or inattentive when with a female teacher.[6] 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, the estimated effects of a female teacher on *NOTLF, NOTUSE*, and *AFASK* are, for girls, statistically indistinguishable from zero.

But the results in Table 8 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 disruptive or inattentive. And, when assigned to a female science teacher, girls are significantly less likely to claim that they do not look forward to science or that science is not useful for their future. The heterogeneity across subjects is only statistically significant in the case of the *NOTUSE* variable. Nonetheless, when combined with the results in Table 6, the results in Table 8 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 9 presents the key results of similarly specified models when applied to the data on boys. Some of the estimates in Table 9 are statistically imprecise. However,

5. Interestingly, because some teachers have sampled students of different genders but the same classrooms, it is also possible to introduce fixed effects for each of the more than 5,400 unique classrooms. And doing so leads to a similar point estimate (that is, -0.035). However, unsurprisingly, this approach also exhausts most of the sample variation associated with the 8,814 first-differenced observations and leads to a dramatically increased standard error (that is, 0.029).

6. However, female teachers are more likely to view both boys and girls as not completing their homework, a finding that may reflect gender differences in the assignment or expectations about homework.

Table 8
FD Estimates of the Effect of a Female Teacher on Girls' Nontest Outcomes, Excluding Mathematics

Dependent variable	Female teacher	By academic subject: Science	English	History	p-value (H_0: $\beta_S = \beta_E = \beta_H$)
DISRUPT	−0.021**	−0.028**	−0.004	−0.021	0.7124
	(0.010)	(0.012)	(0.026)	(0.016)	
INATT	−0.032**	−0.042**	−0.006	−0.030	0.6399
	(0.013)	(0.018)	(0.032)	(0.024)	
NOHWK	0.024**	0.027*	0.025	0.016	0.9084
	(0.011)	(0.015)	(0.024)	(0.021)	
NOTLF	−0.066	−0.111**	−0.006	−0.020	0.4796
	(0.042)	(0.055)	(0.099)	(0.080)	
NOTUSE	−0.037	−0.108***	0.076	0.024	0.0282
	(0.031)	(0.042)	(0.071)	(0.051)	
AFASK	−0.007	0.008	−0.017	−0.034	0.8151
	(0.030)	(0.040)	(0.060)	(0.057)	

Note: 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; and *** Statistically significant at the 1 percent level.

they do indicate that boys are significantly more likely to be seen as disruptive when assigned to a female teacher.[7] Furthermore, boys are significantly more likely to report that they do 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 the effects associated with a female teacher are particularly pronounced in history and English classes.

VIII. Conclusion

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. Furthermore, the sizes of the estimated effects

7. 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 (for example, as in Table 7). The estimate from that approach (0.023 with a *p*-value of 0.005) is quite similar to those reported here, indicating that the results in Tables 8 and 9 cannot be attributed to gender patterns in teacher unobservables.

Table 9
FD Estimates of the Effect of a Female Teacher on Boys' Nontest Outcomes, Excluding Mathematics

Dependent variable	Female Teacher	By academic subject			p-value (H_0: $\beta_S = \beta_E = \beta_H$)
		Science	English	History	
DISRUPT	0.034**	0.013	0.058**	0.060**	0.1667
	(0.013)	(0.018)	(0.029)	(0.024)	
INATT	0.003	−0.019	0.035	0.024	0.2394
	(0.015)	(0.021)	(0.033)	(0.027)	
NOHWK	0.031*	0.020	0.016	0.065**	0.3205
	(0.016)	(0.020)	(0.035)	(0.027)	
NOTLF	0.095**	0.020	0.098	0.250***	0.0147
	(0.043)	(0.051)	(0.110)	(0.067)	
NOTUSE	0.009	−0.050	0.124	0.042	0.1023
	(0.034)	(0.042)	(0.085)	(0.058)	
AFASK	0.042	0.029	0.031	0.077	0.6423
	(0.030)	(0.037)	(0.053)	(0.051)	

Note: 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; and *** Statistically significant at the 1 percent level.

are quite large relative to the subject-specific gender gaps. For example, assignment to an opposite-gender teacher lowers student achievement by nearly 0.05 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 and would do so by improving the performance of boys and simultaneously harming the performance of girls. More specifically, changing an English teacher from female to male would lower the achievement of girls by 0.045 standard deviations and raise the achievement of boys by 0.047 standard deviations (Table 6). The resulting reduction in the gender gap (0.092 standard deviations) is roughly a third of the current difference among 13-year-olds (Table 1). Similar calculations suggest that switching from a male to a female teacher would close the gender gap in science achievement among 13-year-olds by more than half and eliminate entirely the smaller achievement gap in mathematics (Table 1).

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 disruptive. However, the estimates presented here indicate that a year with a male teacher would close the gender gap in the probability of being seen as disruptive by half (Tables 8 and 9). Furthermore, the variable indicating whether a student does not view science as useful for their future is 0.20 standard deviations higher for girls than for boys. However, the estimates presented here imply that a year with a female science teacher would close the gender gap in seeing science as useful for one's future by at least half (Table 8).[8]

The sizes of the estimates reported here suggest that the gender dynamics between teachers and middle-school students have a substantial influence on the gender differences in several important educational outcomes. However, the degree to which the gender dynamics within classrooms contribute to the observed gender gaps in specific subjects (for example, Table 1) depends critically on the gender distribution of teachers by subject. Calculations based on data from the U.S. Department of Education's 1999–2000 Schools and Staffing Survey (SASS) indicate that the percentage of eighth grade teachers who were female ranged from 48 percent in history to 83 percent in English. In math and science, the percentage of female teachers was 65 and 58 percent respectively. When combined with the test score results, the SASS results 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, the test-score results in Table 7 imply that, if half of the reading teachers in sixth, seventh, and eighth grades were male (and their effects were additive), the gender gap in reading skills would fall by approximately a third. Similarly, the results in Tables 8 and 9 suggest that part of boys' relative propensity to be seen as disruptive is due to the gender interactions associated with 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.

It should be clearly noted that the direct point of the simple calculations discussed above is not to suggest that the gender-based segregation of students and teachers would be a desirable policy. Instead, it is to suggest that the results presented here indicate that the gender interactions between students and teachers constitute a quantitatively important environmental determinant of the comparative educational outcomes of both girls and boys. One interesting policy aspect of this general finding involves its implications for the new initiatives designed to reward teachers for the value added to their student's achievement. In particular, the results in this study suggest that such initiatives could penalize teachers for consequences that are unrelated to their behavior (for example, reductions in the achievement of opposite-gender students due to stereotype threat or role-model effects). However, value-added incentives might also encourage teachers to remediate the pejorative achievement consequences associated with gender interactions.

The implications of this study's results for policy efforts to promote gender equity are not clear and would turn critically on understanding more about the structural nature of interactions within classrooms. In particular, the results presented in this study do not speak to the likely effects of proposals for single-sex schooling, which would involve changing the gender distribution of students (not teachers) and training teachers of both genders in gender-specific teaching methods (Sax 2005). Similarly, this study's results do not identify the likely consequences of segregating students and teachers by gender, which would change the composition of student peers and raise a variety of other moral and practical concerns. Instead, what the

8. One caveat regarding the effect sizes is that they may overstate the effect of a single-year's assignment to the extent that individual teachers remained with a cohort of students over several years.

reduced-form results presented here do suggest is that the gender interactions between students and teachers are consequential and that it would be worthwhile to know more about why such student-teacher interactions matter. More specifically, understanding the relative contributions of the various hypotheses consistent with the results presented here (for example, role-model effects, stereotype threat, and teacher biases) would provide a useful guide for developing appropriately targeted policies that shape the gender patterns of educational outcomes in normatively desirable ways.

References

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

Ashenfelter, Orley, and Alan Krueger. 1994. "Estimating the Returns to Schooling Using a New Sample of Twins." *American Economic Review* 84(5):1157–73.

Ashenfelter, Orley, and Cecilia Rouse. 1998. "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics* 113(1):253–84.

Bettinger, Eric P., and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review* 95(2):152–57.

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

Cahill, Larry. 2005. "His Brain, Her Brain." Scientific American 292(5):40–47.

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

Cavanagh, Sean. 2005. "Educators Revisit Girls' Loss of Math, Science Interest." *Education Week*, May 4, p. 6.

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

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

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

Dee, Thomas S., and Sarah R. Cohodes. Forthcoming. "Out-of-field Teachers and Student Achievement: Evidence from Matched-pairs Comparisons" *Public Finance Review.*

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(3):547–61.

Freeman, Catherine E. 2004. *Trends in Educational Equity of Girls and Women: 2004.* NCES 2005-016, U.S. Department of Education, National Center for Education Statistics. Washington, D.C.: GPO.

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(2):129–46.

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

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "The Market for Teacher Quality." NBER Working Paper No. 11154, Cambridge, Mass.: National Bureau of Economic Research.

Hedges, Larry V., and Nowell, Amy. 1995. "Sex Differences in Mental Test Scores, Variability, and Numbers of High-scoring Individuals." *Science* 269(5520):41–45.

Hoffer, Thomas B. 1992. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." *Educational Evaluation and Policy Analysis* 14(3):205–27.

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

Jacob, Brian A. 2002. "Where the Boys Aren't: Noncognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21(6):589–98.

Jones, Susanne M. and Kathryn Dindia. 2004. "A Meta-analytic Perspective on Sex Equity in the Classroom." *Review of Educational Research* 74(4):443–71.

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

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

Lavy, Victor. 2004. "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment." NBER Working Paper No. 10678, Cambridge, Mass.: National Bureau of Economic Research.

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

Lippa, Richard A. 2002. *Gender, Nature, and Nurture*. Mahwah, N.J.: Lawrence Erlbaum.

Loveless, Tom. 1998. "The Tracking and Ability Grouping Debate." *Thomas B. Fordham Foundation Publications*. July 1998. Retrieved September 13, 2006 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=127&pubsubid=804.

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. 1996. "Intelligence: Knowns and Unknowns." American Psychologist 51(2):77–101.

Neumark David, and Rosella Gardecki. 1998. "Women Helping Women? Role Model and Mentoring Effects on Female PhD Students in Economics." *Journal of Human Resources* 33(1):220–46.

Nixon, Lucia A., and Michael D. Robinson. 1999. "The Educational Attainment of Young Women: Role Model Effects of Female High School Faculty." *Demography* 36(2):185–94.

Riordan, Cornelius. 1999. "The Silent Gender Gap." *Education Week*, November 17, p. 46, 49.

Robst, John, Jack Keil, and Dean Russo. 1998. "The Effect of Gender Composition of Faculty on Student Retention." *Economics of Education Review* 17(4):429–39.

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

Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review* 48(3):515–30.

Rouse, Cecilia. 1999. "Further Estimates of the Economic Return to Schooling from a New Sample of Twins." *Economics of Education Review* 18(2):149–57.

Sadker, Myra, and David Sadker. 1994. *Failing at Fairness: How Our Schools Cheat Girls*. New York: Touchstone.

Sadker, David. 2002. "An Educator's Primer to the Gender War." Phi Delta Kappan 84(3):235–44.

Saltzman, Amy. 1994. "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*, November 6, page 4.

Sax, Leonard S. 2005. "The Promise and Peril of Single-Sex Public Education." Education Week, March 2, p. 48, 34–35.

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

Spencer, Bruce D., Martin R. Frankel, Steven J. Ingels, Kenneth A. Rasinksi, and Roger Tourangeau. 1990. *National Education Longitudinal Study of 1988 Base Year Sample Design Report*. NCES 90–463, Washington, D.C.: U.S. Department of Education.

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

U.S. Department of Education. 2000. *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. NCES 2000–469, by Jay R. Campbell, Catherine M. Hombo, and John Mazzeo. Washington, D.C.: Office of Educational Research and Improvement, National Center for Education Statistics.

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