Looking Beyond Academic Performance:

The Influence of Instructor Gender on Student Engagement and Attitude in STEM Fields

Sabrina Solanki (Correspondence author)
University of California, Irvine
Education 3225, Irvine, CA 92697
ssolanki@uci.edu

Di Xu

University of California, Irvine

Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant Number 1256500. The opinions expressed are those of the authors and do not represent views of the National Science Foundation. We would like to thank Mark Warschauer, Lynn Reimer, and Amanda Nili for providing high-quality data and expert guidance on the course context of this research. We are also indebted to Greg Duncan, George Farkas, Rachel Baker, Emily Penner, and Phoebe Camilletti for their valuable comments and suggestions on drafts of this paper.

Abstract

Recruiting more female college faculty has been suggested as a policy option for addressing gender disparities in STEM fields by better engaging female students through a "role model effect". While a small but growing body of literature has examined the role of instructor gender at the higher education level, it typically only focuses on academic outcomes. As a result, the extent to which instructor gender influences student-instructor interactions, engagement, and attitude remains largely unknown. This paper exploits a unique dataset that not only includes student course performance in STEM, but also self-reported non-cognitive measures, namely engagement, self-efficacy, interest, and utility value. We find that having a female instructor narrows the gender gap in successful course completion, behavioral engagement, and interest.

Keywords: gender match STEM instructor gender non-cognitive engagement

1. Introduction

Given the demands of modern society, STEM has an indisputably prevalent and significant role in higher education. However, the choice of a STEM major in college remains highly gender-dependent. Roughly 14.5% of female college students major in a STEM field--as compared to 32.9% of males--and 32.4% of college-going females leave STEM to study unrelated fields, whereas 25.5% of males do so (Chen, 2009). It is not surprising, then, that women comprise a small proportion of fields like computer science, engineering, and physics (NSF, Science and Engineering Indicators, 2016), which generally confer high earnings (BLS, 2013). As a result, gender gaps in STEM education translate to disproportionate gains in long-term economic mobility, earnings inequality, and occupational segregation when students enter the labor market. While a variety of policies have been implemented to recruit more college students from underrepresented groups into STEM, the question of persistence is left unanswered; how do female college students remain in and best progress through the STEM pipeline? To address this question, this study examines the impact of instructor gender on student engagement and attitude in STEM fields.

Recent literature suggests that female STEM success may be strongly influenced by students' educational experiences in college (Chang, Sharkness, Hurtado, and Newman, 2014). Theoretical discussions have nominated a few reasons why gender may be relevant in this context. Among them is the idea that a student-instructor gender match has implications for the identity development process, as female instructors can serve as academic role models (Lockwood & Kunda, 1997). A female instructor may help female students shape and maintain a STEM-related identity (Gilmartin, Denson, Li, Bryant, Aschbacher, 2007), become more engaged during the learning process (Schunk, 1995), and become confident that their future in STEM is attainable (Oyserman, 2007).

A small but growing number of studies have examined the role of instructor-student gender match in college classrooms (e.g., Canes & Rosen, 1995; Neumark & Gardecki, 1998; Robst, Keil, & Russo, 1998; Bettinger & Long, 2005; Hoffman & Oreopoulos, 2009). With the exception of Carrell, Page, & West (2010), who found substantial same gender effects, these studies generally found small positive effects of having a same-gendered instructor for academic indicators of success such as course performance. Yet, due to data limitation, none of these studies were able to explore much beyond these types of academic performance indicators. While academic measures are important, a rich body of literature shows that a more inviting learning context might benefit students by improving their motivation, attitude, and engagement without necessarily boosting students' immediate course performance. Further, improved motivation, engagement, and attitude may have an influential impact on students' later choices and success (Karabenick & Urdan, 2014), such as future educational and employment outcomes (Gutman & Schoon, 2013). The role of instructor-student gender match therefore cannot be fully understood without directly exploring students' course engagement and attitude.

Drawing on a unique dataset that combines college-level student achievement and demographic data with self-reported academic engagement outcomes, this study examines the impact of instructor gender on student course engagement and attitude in STEM gateway courses. We choose to focus on gateway courses because the students' learning experience in these courses often affect their interest and success in subsequent STEM learning, also influencing important academic decisions such as major choice and even college persistence (Chambliss & Takacs, 2014).

Our focus on non-academic outcomes, including both engagement (such as class attendance, participation, and interpersonal interaction with the instructor) and attitude (such as self-efficacy, confidence, and interest), fills an important gap in the current literature that mainly uses academic achievement as the key outcome measure. Not only are these non-academic outcomes important measures of student experience and success themselves, but they also shed light on the channels through which the instructor-student gender interaction may influence the learning process and experience and whether students persist when facing challenges. Indeed, the larger constructs represented by our engagement and attitude measures have been associated with academic persistence and later adult outcomes (Gutman & Schoon, 2013).

The major methodological challenge in estimating the impact of a same-gendered instructor on students' learning outcomes is student self-selection into courses taught by male and female instructors. To minimize selection bias, we build on existing literature by using a two-way fixed effects model (e.g., Hoffman & Oreopoulos, 2009; Fairlie, Hoffman, & Oreopoulos, 2014) that controls for both student-level fixed effects and class fixed effects (i.e., a course-section level), therefore eliminating potential bias that is constant at either the individual level (such as academic capacity) or at the classroom level (such as instructional quality and grading criteria).

We begin our inquiry by examining whether introductory STEM courses with female instructors, as compared to male instructors, narrows the achievement gap between genders, as measured by course grades and probability of successfully completing the course with a C or above. As detailed below, our fixed effects analysis yields results similar to those of previous studies: female students, on average, are less likely to successfully complete a STEM gateway course than male students, but the gender achievement gap shrinks in the presence of a female instructor. These results then motivate the main research question of this study: does having a same-gendered instructor influence the gender gap in course engagement and attitude measures? To address this question, we utilize self-reported student pre-and post-course survey data measuring four important non-cognitive measures, including student course behavioral engagement, self-efficacy in a particular field, interest in this field, and utility value of this field. Although we find that female students tend to be less engaged and interested than male students, overall, such gender gaps in course behavioral engagement and attitude decrease when a female instructor teaches a STEM course.

In the following section, we introduce a theoretical framework and discuss relevant studies linking instructor gender to student academic performance and motivation. Next, we describe our data and research context, sample, and methodology, eventually presenting our results and concluding with a discussion of both our findings and their implications for policy and practice.

2. Theoretical Discussion and Empirical Evidence for Student-Instructor Gender Match

Existing literature exploring the impact of gender on student learning outcomes has built on three major explanations as to how, for example, a female instructor might influence female students' academic performance and engagement. These explanations include role model effects, differences in teaching style, and teacher bias. Below we discuss these three explanations sequentially, review empirical evidence in support of each of these theories, and then discuss implications for our study.

2.1 Identity-based Motivation Theory

The university experience is a crucial setting in which to study student development and the influence of factors such as instructor gender.[1] During this developmental time, students begin to shape their identity (Erikson, 1968; Waterman, 1985), and imagine their future selves while making associations between the college classroom and the larger society it represents. The idea that instructors' gender can contribute to a role model effect for college students is rooted in identity-based motivation theory, which posits that identities are not stable, but dynamically constructed within contexts that help individuals interpret experiences and react accordingly (Oyserman, 2007). An academic interpretation can either be effort-undermining (e.g., decreased level of engagement during the learning process) or effort-enhancing.

Individuals are most likely to pursue opportunities that are identity-congruent, so it follows that female role models can help female students interpret experiences in a positive way by making them more identity-congruent. Prior studies have shown that improved identity compatibility is directly linked to better academic performance and an increased motivation for females (see London et al. 2011, Rosenthal et al. 2011, and Settles et al. 2009). Thus, if more STEM courses were taught by female professors, female students could experience the identity-congruent benefits of being more engaged in STEM, exerting more effort to overcome academic challenges, and persisting more in moving toward STEM-related goals.

Research also shows that female instructors protect against negative consequences that arise from *stereotype threat*, or decreased performance resulting from the fear of confirming a negative stereotype (Steele, 1997). This is an important finding given that the stigma of females in STEM fields is often responsible for female STEM attrition (AAUW, 2008). Indeed, Marx & Roman (2002) observed that female students performed more poorly than male students (and exhibited lower self-esteem) when a math test was administered by a man, but performed on par with male students when the test was administered by a woman. Moreover, McIntyre et al. (2005) found that when females read essays about successful women, this alleviated performance deficits that may have resulted from stereotype threat. Overall, the theoretical discussions suggest that the psychological stress undermining female academic performance in STEM-related fields and lowering self-efficacy fades in the presence of a strong female role model.

Studies indirectly testing identity-based motivation theory and role model effects have been conducted extensively in the K-12 context, and results vary (e.g., Dee, 2005; Nixon and Robinson, 1999; Ehrenberg, Goldhaber, and Brewer, 1995; Lavy, 2004; Zirkel, 2002; and Antecol et al. 2015). For example, Dee's (2005) seminal work found that middle-school aged girls were more likely to report that they did not look forward to science and/or did not think science would be useful for their future when assigned to a female science teacher. On the other hand, Zirkel (2002) found that middle school-aged students with at least one race-and gender-matched role model had better long-term academic outcomes, reported more achievement-oriented goals, enjoyed achievement-related activities to a greater degree, and thought more about their future when compared to students without a race-and gender-matched role model.

Given a push to diversify faculty within college departments, a small but growing body of literature has focused on instructor-student gender interaction at the postsecondary level (e.g., Canes & Rosen, 1995; Neumark & Gardecki, 1998; Robst, Keil, & Russo, 1998; Bettinger & Long, 2005; Hoffman & Oreopoulos, 2009; Carrell, Page, & West, 2010). In terms of academic achievement measures, results are mixed, with some finding a positive impact for initial course performance (e.g. Carrell et al., 2010) and others finding a null effect (Hoffmann & Oreopoulos, 2009). Unfortunately, due to data limitations, the majority of gender match studies could only focus on academic performance and do not have direct measures of behavioral engagement and attitude. Thus, the mechanisms behind student academic success are, at this point, ambiguous.

2.2 Teaching Styles

Outside of role model effects, instructional technique can also explain the impact of student-instructor interaction on student achievement and engagement. Indeed, prior research has shown that no single approach to teaching works for every student, particularly in STEM instruction (Carrell et al. 2010), and that teaching preferences do vary by student gender. For example, in one study, researchers found that male college students in a STEM lab course preferred multimodal instruction, whereas most female students in the class preferred single-mode instruction utilizing a “learning by doing” approach (Wehrwein, Lujan, & DiCarlo, 2007). Learning preferences have consequences for student learning, as students are more likely to be engaged in courses that employ their preferred instructional techniques, and their engagement can lead to improved performance (Ferrara, 2012).

While we do know that student gender matters in terms of preferred instructional approach, less is known about whether teaching style correlates with instructor gender. In one study, female instructors were found to often use interactive teaching techniques such as class discussion, small group discussion, and group projects. Male instructors, on the other hand, used less interactive approaches, tending to rely more heavily on lecture (Starbuck, 2003). Current evidence suggests that gender differences like these do have implications for student learning; specifically, female students prefer certain types of teaching styles that are associated with instructor gender (Carrell et al., 2010).

2.3 Teacher Bias

Another student-instructor factor that might influence student learning and engagement outcome is teacher bias, whereby it is possible that teachers treat their students differently based on gender. Literature has mainly looked at teacher bias through the lens of expectations, often citing the Pygmalion effect, or the notion that people perform better when more is expected of them (Rosenthal & Jacobsen, 1968); it is logical to conclude that gender stereotypes are highly relevant in the present study. For example, boys are often viewed as being better at math, and thus STEM teachers might have higher expectations for them, interact more with them, and overestimate their performance. These actions could improve male achievement and engagement while at the same time creating challenges for females in STEM disciplines (see Rebhorn & Miles 1999; Sadker & Sadker 1985; Hyde & Jaffee 1998). Additionally, in a recent study, Lavy & Sand (2015) compared the scores of middle-schoolers' math-tests, which were graded by the students' math teachers, to scores of tests graded by external teachers who were therefore completely unaware of student gender. The students' teachers tended to give female students lower scores and male students higher scores than did the external group of teachers.

The possible existence of teacher bias explains why test scores alone are insufficient in understanding the experience of students in STEM courses and the impact of instructor gender. For example, if a teacher is more likely to give higher grades to a student who shares his/her gender, a positive estimate of student-instructor gender match regarding test scores would not translate into a positive learning experience for female students; it would instead simply reflect teachers' response to students' gender in assigning a grade. Existing studies have addressed teacher grading bias by exploring longer-term outcomes, such as students' subsequent enrollment and performance in STEM courses and their likelihood of graduating with a STEM degree (e.g. Carrell et al., 2010). Our study adds new evidence to existing literature by directly examining measures of course engagement and attitude.

2.4 Implications for the Current Study

Overall, there is limited evidence about the effect of student-instructor gender match at the higher education level. Even among studies that identify a positive impact on student course performance, however, researchers fail to reach a consensus on the mechanisms driving these results. One important reason for conflicting interpretations is the lack of ability to look beyond achievement scores and directly measure student behavior, engagement, and motivation.

Our study builds on previous studies by providing a current perspective on the educational relevance of gender in a college setting that is congruent with the experiences of many U.S. college students. Most importantly, it fills a gap in educational literature with a unique examination of the influence of same-gendered instructors on non-academic outcomes, using pre-and post-self-reported engagement and attitude variables as key dependent variables. To the knowledge of the authors, a thorough examination of the influence of instructor gender on student motivation and engagement has been unstudied in prior literature, thereby substantiating the purpose of this study.

3. Study Background

3.1 Data and Research Context

Our analysis is based on detailed student and administrative data from the University of California, Irvine (UCI), a large public university in California. UCI is part of the University of California system, which consists of ten campuses with a total enrollment of approximately 200,000 undergraduate students in the 2013 school year. UCI enrolled approximately 12.5% of these students.

The UCI student population consists of approximately 45% Asians and 30% Caucasians, with roughly a quarter of the student body belonging to an under-represented minority group. Of its 25,000 undergraduate students in the 2013 school year, more than half were female, and 30% of degrees awarded in 2013 were in a STEM field.

Beginning in the spring of 2013, a team of researchers at UCI collected the dataset used in this paper, observing full and part-time undergraduate university students enrolled in 23 different STEM courses during one year (3 academic quarters). These courses include introductory courses in the field of biological sciences (e.g., 'DNA to Organisms,'), Chemistry (e.g., 'General Chemistry,'), Engineering (e.g., 'Introductory Engineering'), Information and Computer Sciences (e.g., 'Introductory Programming'), Math and Statistics (e.g., 'Introduction to Linear Algebra'), and physical sciences (e.g., 'Classical Physics'). Furthermore, these courses were prerequisites for other mandatory courses in one or more STEM majors and were offered in multiple sections during the year. Each typically enrolled 200 or more students. Table A1 in the Appendix provides descriptive statistics for the list of courses used in this study.

Detailed student-level information, collected from the Office of Institutional Research (OIR), includes demographic information regarding gender, ethnicity, age, first-generation status, and low-income status; academic preparation variables, such as SAT math and verbal score, high school GPA, and AP exam scores; and postsecondary data, such as the type of courses students take during a particular quarter, the number of credits students take in a quarter, cumulative and current GPA, and programs of study. Student administrative records are then matched to courses in the dataset, which includes information about instructor name, gender, and academic rank.

3.2 Engagement Measures

During the year of data collection, pre- and post-experience surveys were administered in students' STEM courses. These surveys were administered during week 2 and week 9 of the 11-week quarter and were completed by approximately 55% of students in our sample.[2] The survey items represent different motivational constructs such as interest value, utility value, and self-efficacy; they also identify students' level of course engagement, as measured by attending class, listening attentively to lectures, engaging in formal help-seeking behavior, and participating in class discussion. These items were standardized by course and used in our analysis as outcome variables.

Specifically, the survey included 25 items to measure students' experience in the course and motivation toward the course subject, where 12 items were not relevant to our study and are therefore not included.[3] The items used in our study were divided into 4 major categories: behavioral engagement, academic self-efficacy, interest, and utility value. We include details for each category and a Cronbach's internal reliability alpha score below. This score is an index of internal consistency reliability that assesses the degree to which responses are consistent across a set of multiple measures of the same construct, usually self-report items (Warner, 2013). Below we introduce the scale of each variable used in our study.

Behavioral engagement: This category includes four variables in the post-experience surveys only. Specifically, students were asked, "During this course, about how often have you done the following respectively: (i) attended lectures, (ii) listened attentively to lectures, (iii) asked the professor or TA for help in this class, and (iv) asked questions and contributed to the discussion in lecture?" Students were asked to indicate their response on a 5-point Likert scale, with 1 indicating *never*, 2 *rarely*, 3 *sometimes,* 4 *often,* and 5 *always*. The Cronbach's internal reliability alpha is equal to 0.62.

Academic self-efficacy: For the pre- and post-experience surveys, students were asked to respond to the following three items using a 5-point Likert scale ranging from 1 indicating *not at all true* to 5 indicating *very true*: (i) "I'm certain I can master the skills taught in this class," (ii) "I'm certain I can figure out how to do the most difficult course material," and (iii) "I can do almost all the work in the class if I don't give up." The Cronbach's internal reliability alpha is equal to 0.86.

Interest: For the pre- and post-surveys, students were asked to respond to the following three items using a 5-point Likert scale ranging from 1 indicating *not at all true* to 5 indicating *very true*: (i) "I find many topics in this course to be interesting," (ii) "Solving problems in this class is interesting for me," and (iii) "I find this class intellectually stimulating." The Cronbach's internal reliability alpha is equal to 0.89.

Utility value: For the post-experience survey, students were asked to respond to the following two items using a 5-point Likert scale ranging from 1 indicating *not at all true* to 5 indicating *very true*: (i) "Doing well in this course is important to me" and (ii) "Understanding the material in this course has many benefits for me." Cronbach's internal reliability alpha is equal to 0.60.

3.3 Sample and Summary Statistics

The analytic sample consists of 20,429 course observations taken by 9,766 students who were enrolled in STEM courses.[4] Table 1 presents summary descriptive statistics at the student-class level. Compared to the general UCI student population described in section 3.1, students in our analytical sample consist of more Asian students, a smaller proportion of students from under-represented ethnic groups, and a slightly smaller proportion of females.

Table 1 also documents differences, sorted by gender, in our two main academic outcomes: average course grade and receiving a grade of "C" or higher in a course. We chose

"C" or higher as a threshold mainly because, at most universities, dropping below a "C" average places students on academic probation, which is a strong predictor of college attrition (Seymour & Hewitt, 1997). Grades are determined on a 4-point scale whereby an A is 4 points, an A- is 3.7 points, a B+ is 3.3 points, and so on. An F is 0 points. The gender achievement gap remains consistent irrespective of instructor gender. However, male students seem to be subject to a performance decrement when switching from a male instructor (M_{male}=2.67) to a female instructor (M_{female}=2.57) while the average performance of female students seems to be fairly stable (M_{male}=2.48 versus M_{female}=2.47). Also, both male and female students are more likely to receive at least an average grade ("C" or higher) with male instructors than they are with female instructors. In all subsequent regression analyses, course grades are standardized by course so that grade changes can be understood as effect sizes.

Table 1 also documents differences, sorted by gender, for our 13 engagement variables. All variables have been standardized to have a mean of zero and a standard deviation of one within a course. The descriptive evidence provided in Table 1 shows male students reporting higher levels of engagement than female students across all measures except for utility value, for which female students report higher values. The difference between male and female students' reported values is particularly pronounced for the items measuring self-efficacy. In fact, for the item "I'm certain I can figure out how to do the most difficult course material," the raw difference is 0.35 standard deviation units. The gender difference in reported values for the items measuring interest ranges from 0.08 to 0.18 standard deviation units.

Table 2 presents summary descriptive statistics for our STEM courses. Each student, regardless of gender, took an average of 2.1 courses. 21% of females took classes in biology, whereas 12% of males did so. Roughly 54% of females and 41% of males took classes in chemistry. Percentages for female and male involvement in mathematics and statistics classes are roughly the same. Substantially fewer females took computer science and engineering classes than males. These gender differences in course-taking are similar to what national data indicates (Chen, 2009). There is also a gender disparity in instruction. 39% of instructors across all courses in our dataset are female (34% are female, on average, per course). The dataset includes 73 sections (defined as a course delivered at a specific time for which students meet with an instructor) within 23 courses over 3 academic quarters, with 3 sections on average per course.

4. Estimation Strategy

4.1 Basic Ordinary Least Squares Model

In this study, we ask whether switching from a male instructor to a female instructor improves female students' performance and engagement in STEM-related gateway courses relative to male students (i.e., the gender gap). We begin with a basic ordinary least squares (OLS) model (Equation 1).

$$Y_{ijkst} = \beta_0 + \beta_1(\mathit{Female\_instructor}_j) + \beta_2(\mathit{Female\_student}_i) + \beta_3(\mathit{Female\_instructor}_j * \mathit{Female\_student}_i) + X_i + \mathrm{Rank}_j + \delta_t + u_{ijkst} \qquad (1)$$

Our data is organized at the student-by-course-section level. Y_{ijkst} represents outcome Y for student *i* in course *k* section *s*, with instructor *j* in term (i.e., quarter) *t*. $\mathit{Female\_instructor}_j$ is dichotomous and equal to 1 if instructor *j* is female and 0 if instructor *j* is male. $\mathit{Female\_student}_i$ is dichotomous and equal to 1 if student *i* is female and 0 if student *i* is male. X_i is a vector of student control variables and includes student ethnicity, first-generation status, low-income status, SAT math score, SAT verbal score, and high school GPA. Rank_j is an indicator of academic rank of instructor *j*. δ_t represents term fixed effects and reflects the specific quarter when a student took a course. The interpretation of equation 1 relies on three key estimates. Using standardized course grade as an example, β_1 represents the average changes in standardized course grade for male students with a female instructor, relative to a male instructor. β_2 represents the average changes in standardized course grade for female students, relative to male students (i.e., the gender achievement gap), with a *male* instructor. $\beta_2 + \beta_3$ represents the average changes in standardized course grade for female students, relative to male students (i.e., the gender achievement gap), with a *female* instructor. The interactional effect (β_3) therefore represents the decrease (or increase) in the gender achievement gap and is the main parameter of interest.

4.2 Two-way Fixed Effects Model

Using a basic OLS model to estimate an unbiased interaction effect between female students and instructors is quite challenging given the non-random assignment of students to instructors. For example, we might be concerned that students who prefer taking a course with a female instructor, rather than a male instructor, might have certain characteristics that also correlate to their potential course performance. If any such student-level characteristics (e.g., motivation and ability) are absent from the dataset, the estimates from a basic OLS model would be biased. In this case, a positive interaction effect might be due to student sorting between male and female instructors and not necessarily a same-gender student-instructor match.

Similarly, bias could also arise from unequal distribution of female and male instructors across different types of courses. For example, male instructors might be assigned to teaching more challenging courses, whereas female instructors might be more likely to teach more engaging or less academically demanding courses. In such a case, students may perform better or worse in courses that are more or less interesting to them because of the topic, not the instructor.

To address potential biases at either the student level or at the course level, we take advantage of the panel data structure in our study—multiple observations per student—which enables us to employ a two-way fixed effects model.

$$Y_{ijkst} = \beta_0 + \beta_1(\textit{Female\_instructor}_j) + \beta_2(\textit{Female\_student}_i) + \beta_3 (\textit{Female\_instructor}_j * \textit{Female\_student}_i) + \text{Rank}_j + \gamma_i + \delta_k + \delta_t + \Phi_{kg} + u_{ijkst} \quad (2)$$

Compared to equation 1, equation 2 includes 2 additional sets of variables: γ_i and δ_k, which are fixed effects terms for student and course, respectively. γ_i controls for both observable and unobservable student-level characteristics that are constant for an individual (e.g., ethnicity and academic motivation). This term does not, however, control for how student achievement motivation applies to some courses but not others.[5] δ_k is a course fixed effects term that captures course-level characteristics (e.g., field of study, pre-requisites, course requirements, course topics, and level of difficulty). Therefore, students are compared only within a particular course (e.g. 'Introduction to Programming'). This specification rules out the possibility that the estimated effects of instructor-student gender interaction are driven by variation across courses. Lastly, Φ_{kg} are course by gender fixed effects, which allow gender differences to vary across courses.

4.3 Section Fixed Effects Model

Even with a two-way fixed effects strategy, potential bias still remains, as the course fixed effects term does not account for differences between sections.[6] For example, the classroom environment and student-instructor dynamics in 'General Chemistry' taught in the Fall by female Professor X, may be different from those for 'General Chemistry' taught during the same term by male Professor Y. Even within the same course, female instructors might be different from male instructors in terms of their instructional approaches and grading policies, and these differences might impact student learning and engagement. However, since our key variable of interest is the interaction between student and instructor gender status, we can employ a section fixed effects term to control for variation across sections within a particular course[7]:

$$Y_{ic} = \beta_3\,(\mathit{Female\_instructor}_j * \mathit{Female\_student}_i) + \gamma_i + \Phi_c + \Phi_{kg} + u_{ic} \qquad (3)$$

As shown in equation 3, Φ_c represents section fixed effects that control for any observable or unobservable section-level characteristics such as class size, delivery format, grading policies, term, and time of day. This specification also implicitly controls for instructor fixed effects; this means that instructor-level variables, such as instructor quality and instructional approach, are also controlled for. A further advantage of including a section fixed effects term, as explained in more detail in Hoffman & Oreopoulos (2009), is that it implicitly standardizes testing procedures across student groups we compare since all students are subjected to the same class elements mentioned above. It is worth mentioning that the main effect for instructor variables (i.e., gender and rank) and student gender has been removed due to controlling for section fixed effects and student fixed effects, respectively.

Similar to the interpretation of equation 1, β_3 in equation 3 represents whether switching from a male instructor to a female instructor reduces the gender gap in achievement or engagement. A positive β_3 indicates that this gap shrinks in the presence of a female instructor. As noted in our earlier theoretical discussion, a positive β_3 for student course grade does not uniquely identify a role model effect, defined as a social psychological effect in which, for example, female students are encouraged simply by observing a female instructor. Instead, a significant interaction between student and instructor gender could be the result of any number of mechanisms including not only role model effect but also favoritism or hostility in assessing male or female students. Additionally, other mechanisms driving an interaction effect might include instructors providing different levels of attention, advice, and encouragement to students based on gender. For example, male professors might encourage and elicit participation from male students (e.g., calling on male students more frequently than female students). Or, female instructors might provide different levels of attention and advice during office hours to her students based on gender. Examining course behavioral engagement and attitude measures can shed light on these possible mechanisms.

4.4 Concerns with Survey Data

The major focus of our study is student non-academic outcomes, as measured by items included in pre-course and post-course surveys. A potential concern regarding these survey items, however, is that instructor-student gender interaction may also influence a student's probability of completing the course survey. For example, if more engaging and motivated students are more likely to complete the course survey in classes taught by a same-gendered instructor, a positive interaction effect for student motivation and engagement measures might be due to student sorting between male and female instructors in completing the course survey, thereby not necessarily capturing students' actual engagement and motivation in a course with a same-gendered instructor.

One way to provide insight on the extent of this problem is to directly assess whether there is a student-instructor gender match effect on students' probability of completing the survey. Therefore, we ran a series of regression models to test for this possibility using equations 1 and 2. In these models, completing a survey was coded as 1, and the results are listed in Table A2 in the Appendix. Neither male nor female students were determined to be more likely to complete a survey as a result of taking a class with a same-gendered instructor, and such null effects are consistent across model specifications. These results therefore alleviate concerns that our response rate might be biased due to the presence of a same-gendered instructor.

5. Results

5.1 Student Achievement Outcomes

Table 3 presents our results estimating the impact of female instructors on the gender achievement gap in college-level STEM courses. Column 1 uses the basic OLS model and includes all available student-level control variables; column 2 adds a course fixed effects term. Column 3 replaces student control variables with student fixed effects, and column 4 represents both section fixed effects and student fixed effects. Standard errors are clustered at the course section level.[8]

The results here are consistent with prior work on gender match although our point estimates are not always significant. As shown in Panel A, the main effects of instructor gender (row 1) echo the descriptive information presented earlier. Specifically, male students receive lower grades in courses taught by female instructors than they do in courses taught by male instructors. On average, taking a course with a female instructor reduces the standardized course grades of male students by a statistically significant amount of 0.07 standard deviations (column 2). Prior studies have reported point estimates ranging from 0.04 to 0.08 standard deviations (Carrell et al. 2010, Hoffman & Oreopoulous, 2009). However, once controlling for student fixed effects (column 3), the estimated impact of instructor gender on male students is no longer significant.

The second row of Panel A displays the β_2 coefficients, which, as noted earlier, represent the gender achievement gap when a male instructor is present. We find a consistent and significant gender achievement gap in courses taught by male instructors. In column 2, which controls for both course fixed effects and student-level covariates, the gender achievement gap is equal to 0.24 standard deviation units. On a 0-4 point scale, this approximately represents the change from a "B+" to a "B," for example. However, the interaction term displayed in the third row of Panel A is not significant, indicating that taking a class with a female instructor cannot significantly improve the gender achievement gap on standardized test scores.

One limitation of using standardized course grade as a continuous variable is that it does not fully capture the possible impact of instructor-student gender match for the full distribution of letter grades. Considering that receiving a "C" or better is particularly important for students' academic success and progress (e.g., maintaining a "C" average in college-level courses is necessary for avoiding academic probation; a quarterly GPA that falls below a "C" average can disqualify students from some majors as well), Panel B of Table 3 further presents the results on students' probability of receiving a grade of a "C" or better in a course.[9]

Although small in magnitude, the coefficient of the interaction term is significant in our preferred model (column 4) that controls for both student-level and section fixed effects. Specifically, the estimates indicate that if a male instructor was replaced with a female instructor, the gender achievement gap would decrease by almost 3 percentage points in favor of female students.

Heterogeneous treatment effects

To explore the validity of a common treatment effect across the student sample, we investigate whether the effects of instructor gender differ by student achievement level. In this analysis, we first divide our sample of students into three categories based on total math SAT score: highest-scoring students (with a score greater than or equal to 660), middle-scoring students (with a score greater than or equal to 590), and lowest-scoring students (with a score less than 590). Next, we examine student performance using both equations (2) and (3) for the three different groups of students. The results are presented in Table 4.

Among the three different performance subgroups, we find treatment effects only for the highest-scoring students, and the results for this particular group are consistent with the full sample analysis results shown in Table 3. For example, we find no significant interaction terms (Panel A, row 3) for the outcome measure of standardized course grade. We do, however, find that the achievement gap decreases in female-taught courses in regard to our second achievement outcome, earning a "C" or better (Panel B), as indicated by the positive β_3 coefficient (row 3). It is important to note that the estimated effect for this subgroup of students ($\beta_3 = 0.05$) is almost twice as large as the effect for the average student. It therefore seems that the gender interaction effect observed in Table 3 is driven mainly by students with the highest SAT scores. This outcome is similar to that of Carrel et al. (2010), which found the most significant effects for female students with the strongest pre-existing math skills.

5.2 Student Engagement Outcomes

Our first step in this paper was to estimate the gender achievement gap in courses taught by male and female instructors. While we do not find evidence to suggest that having a same-gendered instructor has any impact in terms of average course grade, we do find that female students are more likely than male students to receive a "C" or better in female-taught courses. We now turn to our main results, examining whether female instructors improve gender gaps for non-academic outcomes such as engagement and attitude.

Table 5 includes estimates for our variables measuring behavioral engagement. It includes 4 survey items measuring whether a student: attended lectures regularly (Panel A), listened to lectures attentively (Panel B), asked the professor or TA for help (Panel C), and asked questions or contributed to classroom discussion (Panel D). Panel E is an index representing all 4 items together. Column 1 includes all available student-level control variables in addition to a course fixed effects term; column 2 uses both section fixed effects and student control variables to represent our preferred model, as it indicates whether replacing a male instructor with a female instructor increases female students' self-reported engagement--as compared to male students' engagement in the *same course section*.

The results are generally consistent across model specifications for each item in Table 5. Interestingly, we find a significant opposite-gender match for the first item, 'attended lectures' (Panel A). As shown in column 2, which controls for section fixed effects, the gap between female and male students' average reported values is equal to 0.37 (row 2) standard deviations in male taught courses with female students reporting higher values. The gap is equal to 0.29 ($\beta_2 + \beta_3$) standards deviations in female taught courses.

As noted earlier, male students reported higher levels of engagement than female students, indicating the presence of a gender engagement gap (Table 1). The negative β_2 coefficients (row 2) in Panels C & D are consistent across both model specifications and confirm the presence of a gender engagement gap in male-taught courses. Further, our results suggest that a gender match can mitigate this gap. For example, in our section fixed effects model (column 2), female students' average reported value for the variable 'ask the professor for help' is 0.21 standard deviations lower than that of male students in a male-taught course section (Panel C, row 2). However, in a female-taught course section, female students' average reported value is only .02 ($\beta_2 + \beta_3$) standard deviations lower than that of their male counterparts. The decrease in the gender gap with a female instructor, as compared to a male instructor, is significant.

It is important to note that we also find a positive association between instructor gender and our last behavioral engagement item, ‘ask questions or contribute to class discussion.’ As indicated by the positive β_3 coefficient in Panel D (row 3), the gender engagement gap decreases with a female instructor. In addition, when we put our 4 behavioral engagement items together, the interaction effect still holds and is statistically significant ($\beta_3 = 0.10$), as shown in Panel E.

Table 6 includes estimates for our variables that measure interest, including 3 survey items regarding whether a student: finds topics in the course interesting (Panel A), finds solving problems in the class interesting (Panel B), and finds the class intellectually stimulating (Panel C). Panel D is an index representing all 3 items together. All regression estimates control for a pre-score. Like in our behavioral engagement findings, male students reported being less interested in female-taught courses than they did in male-taught courses (point estimates range from 0.07 to 0.10 in row 1, Panels A-D).

We also find that the gender interest gap decreases in female-taught courses, as compared to male-taught courses, for one item measuring interest in the course topic (Panel A). The β_3 coefficient reported in column 1 is significant and indicates that replacing a male instructor with a female instructor shrinks the gender interest gap by 0.08 standard deviations; this result holds in our section fixed effects model, reported in column 2. Lastly, the gender interest gap in male-taught courses for the second item (Panel B) suggests that female students report lower values for the item ‘I find solving problems interesting,’ than do male students in the same course section ($\beta_2 = -0.36$ in column 2, row 2). The interaction term, however, indicates that this gap is not significantly different in female-taught courses.

Lastly, Table 7 includes estimates for our variables that measure self-efficacy, which include 3 survey items regarding whether a student feels certain that he/she can: master the skills taught in class (Panel A), figure out how to do the most difficult course material (Panel B), and do almost all the work in the class by not giving up (Panel C). Panel D is an index representing all 3 items together. All regression estimates control for a pre-score. The β_3 coefficient reported in column 2 (Panel B) indicates that replacing a male instructor with a female instructor decreases the gender efficacy gap for the item, 'I feel certain that I can figure out how to do the most difficult course material', by 0.06 standard deviations.[10]

6. Discussion and Conclusion

Women make up 47% of all U.S. employees yet represent only a quarter of jobs in mathematical sciences and 13% of engineers (NSF, Science and Engineering Indicators, 2016). Statistics like these are affected by the dramatic gender gap within STEM, as 81% of university STEM professors are men, for example. This type of disparity suggests that female students lack role models and are less likely than male students to find STEM compatible with their gender identity. A common policy prescription is to therefore increase the number of female faculty in STEM fields so they may inspire and encourage young women to pursue STEM degrees.

In this study, we have discussed the educational relevance of an instructor's gender in regard to student outcomes in college-level STEM courses. Most of the literature related to this topic in higher education is focused solely on achievement outcomes; our study is the first to look beyond academic performance in examining the relationship between an instructor's gender and non-academic student outcomes. Drawing on the identification strategy used in existing studies on this topic, we used a two-way fixed effects model to examine the extent to which instructor-student gender match influences non-academic variables measuring behavioral engagement, self-efficacy, interest, and utility value.

Using a large dataset of college students, we found no effect of instructor-gender match on average standardized course grade, but a significant -- though small -- positive impact of instructor-gender match on students' likelihood of earning a "C" or better. Based on information collected through pre-and post-experience surveys, we further found that student intellectual engagement can be improved as a result of instructor gender, where the gap between female and male students' course engagement and attitude toward a subject is reduced when a course is taught by a female professor.

Specifically, while female students, as compared to male students, report being less inclined to ask a male STEM professor for help (a difference of 0.21 standard deviations), switching to a female instructor almost completely offsets this gender difference in help-seeking. In a female-taught class, female students' reported value is only 0.02 standard deviations lower than that of their male counterparts on this item. This outcome is an example of how students positively modify their behavior in courses taught by professors who share their gender, validating the premise of this study.

Promisingly, the participation gap also improves in female-taught STEM classes. This is particularly noteworthy given that active participation is generally thought to encourage learning, and most research suggests that male students dominate mixed-discussion groups in the classroom and beyond (Krupnick, 1985). Also, we found that same-gendered instructors influence the interest gap in STEM—a particularly important outcome since research suggests that the gender interest gap in STEM is widening (Neuhauser, 2015).

It is also important to note that our study found a significant result for one item measuring self-efficacy, 'I'm certain I can figure out how to do the most difficult course material.' The earlier descriptive evidence (see Table 1) showed that the gender difference was most pronounced for this item. Further, researchers have found that self-efficacy explains some of the discrepancy between gender representation in STEM majors and careers (Bandura et al. 2001; Hackett & Betz 1989; Pajares, 2002; Oakes, 1990). Our results suggest that a female instructor is able to narrow the gender efficacy gap.

Our study indicates that there is more to discover in the college classroom to help us understand female student success in STEM, also indicating that focusing on non-academic outcomes, such as engagement and attitude, is a promising avenue for future research. Motivation factors like these illuminate why gender matters in the classroom and how certain mechanisms, such as role model effects, may promote student success and persistence within STEM fields. Our future research aim will be to connect these motivation factors to persistence measures, such as number of STEM courses taken and major choice.

Even though the current policy climate is focused on promoting women in science, the results that we found for males should not be ignored. Similar to researchers in prior studies (e.g., Hoffman & Oreopoulos, 2009; Dee, 2005), we found a consistent negative male reaction to a female instructor in terms of both academic and engagement outcomes. For example, male students reported lower values when questioned whether they ask a professor for help (Table 5, Panel C, column 1, $\beta_1 = -0.10$)--or contribute to class discussion in female-taught courses--than they do in male-taught courses (Table 5, Panel D, column 1, $\beta_1 = -0.08$). Also, the β_1 coefficients related to student interest (Table 6), indicating whether male student interest increases or declines with a female instructor, are all negative and significant. In other words, male student interest declines when a course is taught by a female instructor. Some researchers suggest that gender stereotypes within the college environment—specifically regarding student perception of female and male instructors—play a role (Miller & Chamberlin, 2000) in propagating this negative reaction. Future studies should dedicate themselves to understanding the variables that advance the negative reaction of males to a female instructor; perhaps having more encounters with competent female instructors can reverse this perception phenomenon.

The college years are a critical time in which students imagine their future selves and take action to reach future career goals. If females continue to be the minority in STEM classes and are not exposed to readily available STEM female role models, it is likely that they will indeed believe STEM fields are inaccessible, causing female STEM involvement to decline further. This study takes an important first step toward understanding the mechanisms that make gender relevant in the classroom. Future research should continue to examine the role of non-academic outcomes in female student success so that researchers and policymakers can identify potential strategies at the college level that can help female students thrive in the STEM environment.

Notes

[1] Instructor race is an equally important factor to examine in terms of student development in the classroom. Unfortunately, we were unable to examine the implications of race due to lack of variation in instructor race.

[2] 65% of students completed the pre-experience survey, 55% of students completed the post-experience survey, and 50% of students completed both surveys. A potential concern regarding these survey items is that the instructor-student gender interaction may also influence a student's probability of completing the course survey. We directly examine this issue and provide empirical evidence against this possibility in the methodology section (section 4.4).

[3] The excluded survey items did not measure motivational construct that could theoretically be influenced by same-gendered student-instructor interactions. For example, "I have to give up a lot to do well in this course" or "In comparison to the rest of your classmates, how do you rate your achievement?"

[4]Among all the students in the sample, 55% took courses in multiple fields, therefore contributing to the student-fixed effects estimator. However, including students who did not take courses in multiple fields does not bias the results, as long as selection bias, if any, is constant within an individual. A smaller degree of within-individual variation would be problematic if it yielded an imprecise estimator; with large sample sizes, however, this is typically less of a concern. In a robustness check, we also limited the sample to students who took courses in

multiple fields in our sample, and the resulting effect sizes and significance levels were almost identical to those reported here. For a detailed discussion of the properties of the fixed-effects estimator and key assumptions underlying fixed effects models using panel data, see Wooldridge (2002).

[5] We formally test the degree of systematic sorting of students into courses by instructor gender. Specifically, we regress student gender on faculty gender for our full set of courses. The results of this analysis are shown in Table A3 in the Appendix. As shown in column 1, we do not find any significant correlation between student and faculty gender even when we add student controls. We also examine whether there are any differences in the types of female students in female-taught courses. As shown in columns 3 & 4, we regress student academic attributes on an indicator variable for whether a student took a course with a female instructor. Female students in female-taught courses, compared to those in male-taught courses, have slightly higher SAT math scores and high school GPAs. This difference, however, is significant only for the latter.

[6]A section is a course ('General Chemistry') taught during a specific term (winter 2014) at a distinct time with a given instructor. We use the term "section" interchangeably with the term "class."

[7] For our motivation and engagement variables, we do not include models with a student fixed effects term since only 17.86% of students completed multiple surveys; it is unlikely that this small fraction of students represents the qualities of most students in the sample. Instead, we use a vector of student controls.

[8] We also estimate our models by two-way clustering standard errors at the student and course level; standard errors increase but do not change our significance levels.

[9] We also ran models in which receiving a grade of "B" or better and "D" or better functioned as outcome variables. These results were not significant.

[10] The results for both items representing utility value were not significant.

References

AAUW. (2008). *Where the girls are: The facts about gender equity in education*. Washington, DC: Author.

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics, 33*, 63-89.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.

Bandura, A., Barbarnaelli, C., Caprara, G.V., and Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72*, 187-206.

Bettinger, E., & Long, B. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review, 95,* 152-157.

BLS. (2013, September 1). *Labor Force Statistics From the CPS.* Retrieved October 30, 2013, from Bureau of Labor Statistics.

Braxton, J. (2006). Faculty professional choices in teaching that foster student success. Washington D.C. National Postsecondary Education Cooperative.

Canes, B. & Rosen, H. (1995). Following in her footsteps? Women's choices of college majors and faculty gender composition. *Industrial and Labor Relations Review, 48*, 486-504.

Carrel, S., Page, M., and West, J. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics, 125*, 1101-1144.

Chambliss, D. & Takacs, C. (2014) *How college works*. Cambridge, MA: Harvard University Press.

Chang, M., Sharkness, J., Hurtado, S., Newman, C., (2014). What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching, 51*, 555-580.

Chen, X. (2009). *Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education.* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington D.C.

Dee, T., (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95,* 158-65.

Ehrenberg, R., Goldhaber, D., and Brewer, D. (1995). Do teachers' race, gender and ethnicity matter? Evidence from the national educational longitudinal study of 1988. *Industrial and Labor Relations Review, 48*, 547-61.

Erikson, E. (1968). *Identity, youth, and crisis*. New York, NY: Norton.

Fairlie, R., Hoffman, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review, 104*, 2567-91.

Ferrara, M. (2012). The relationship of teaching style, gender preference, and gender match in single gender and coeducational settings. *Advances in Gender and Education, Montgomery Center for Research in Child & Adolescent Development (MCRCAD), 3*, 14-27.

Gilmartin, S., Denson, N., Li, E., Bryant, A., Aschbacher, P. (2007) Gender ratios in high school departments: The effect of percent female faculty on multiple dimensions of students' science identities. *Journal of Research in Science Teaching, 44*, 980-1009.

Gutman, L. & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. London: Institute of Education.

Hackett, G. & Betz, N. (1989). An exploration of the mathematics of self-efficacy and mathematics performance correspondence. *Journal for Research in Mathematics Education, 20,* 261-273.

Hoffman, F., & Oreopoulos, P. (2009). A professor like me: The influence of instructor gender on university achievement. *Journal of Human Resources, 44,* 479-94.

Hyde, S. J., & Jaffee, S. (1998). Perspectives from social and feminist psychology. *Educational Researcher, 27*, 14-16.

Karabenick, S.A. & Urdan, T.C. (2014). *Motivational Interventions*. UK: Emerald Group Publishing.

Krupnick, C. (1985). Women and men in the classroom: Inequality and its remedies. *On Teaching and Learning: The Journal of the Harvard-Danforth Center, 1*, 19-25.

Lavy, V. (2004). Do gender stereotypes reduce girls' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics, 92*, 2083-2105.

Lavy, V. & Sand, E. (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases. (NBER Working Paper No. 20909). Cambridge, MA: National Bureau of Economic Research.

Lewin, T. (2006, July 9). *At Colleges, Women are Leaving Men in the Dust.* Retrieved August 21, 2013, from www.nytimes.com

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73,* 91–103.

Lockwood, P. (2006). Someone like me can be successful: Do college students need same-gender role models? *Psychology of Women Quarterly, 30,* 36-46.

London, B., Rosenthal, L., & Gonzalez, A. (2011). Using experience sampling methodology to capture the impact of identity, support, and gender rejection on the academic engagement of women. *Journal of Social issues, 67*, 510-530.

Marx D. & Roman, J. (2002). Female role models: Protecting women's math test performance. *Society for Personality and Social Psychology, 28*, 1183-1193.

McIntyre, R., Lord, C., Gresky, D., Ten Eyck, L., Frye, G., Bond, C. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology, 10,* 116-136.

Miller, J., & Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology, 28,* 283-298.

National Science Board. 2016. *Science and Engineering Indicators 2016.* Arlington, VA: National Science Foundation (NSB-2016-1).

National Science Foundation. (2011). *Women, minorities, and persons with disabilities in science and engineering: 2011.* Arlington, VA: National Science Foundation.

NCES, N. C. (2005). *Postsecondary Participation Rates by Sex and Race/Ethnicity: 1974-2003.* U.S. Department of Education. Washington, D.C.: Institute of Education Sciences.

Neuhauser, A. (2015, June 29). *Parents key in attracting girls to STEM.* Retrieved August 11, 2016, from www.usnews.com

Neumark D. & Gardecki, R. (1998) Women helping women? Role model and mentoring effects on female PhD students in economics. *Journal of Human Resources, 33*, 220-46.

Nixon, L. & Robinson, M. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography, 36*, 185-94.

Oakes, J. (1990). Opportunities, achievement, and choice: Women and minority students in science and mathematics. *Review of Research in Education, 16,* 153-222.

Oyserman, D. (2007). Social identity and self-regulation. In A. Kruglanski & T. Higgins (Eds.), Handbook of social psychology (2nd ed., pp. 432–453). NY: Guilford Press.

Oyserman, D. (2014). Identity-based motivation: Core processes and intervention examples. In S.A. Karabenick & Urdan, T. (Eds.), *Motivational Interventions* (pp. 213-240). UK: Emerald Group Publishing.

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory into Practice, 41,* 116-125.

President's Council of Advisors on Science and Technology (PCAST). (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: Author.

Rebhorn, L., & Miles, D. (1999). High-stakes testing: Barrier to gifted girls in mathematics and science? *Social Science and Mathematics, 99,* 313-319.

Robst, J., Kell. J., & Russo, D. (1998). The effect of gender composition of faculty on student retention. *Economics of Education Review, 29,* 429-439.

Rosenthal, R., & Jacobsen, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* New York: Holt, Rinehart and Winston.

Rosenthal, L., London, B., Levy, S., and Lobel, M. (2011). The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university, *Sex Roles, 65*, 725-736.

Sadker, M., & Sadker, D. (1985). Sexism in the classroom. *Vocational Education Journal, 60,* 30-32.

Schunk, D.H. (1995). Self-efficacy and education and instruction. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 281-303). New York: Plenum Press.

Settles, I.H., Jellison, W.A., & Pratt-Hyatt, J.S. (2009). Identification with multiple social groups: The moderating role of identity change over time among women scientists. *Journal of Research in Personality, 43*, 856–867.

Seymour, E. & Hewitt, N. (1997). *Talking about leaving: Why undergraduates leave the sciences.* Colorado: Westview Press.

Starbuck, G. (2003, April). *College teaching styles by gender*. Paper prepared for the Western Social Science Association Annual Meeting. Las Vegas, Nevada.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613-629.

Warner, R. (2013). Applied statistics: From bivariate through multivariate techniques. USA: Sage Publications.

Waterman, A. (1985). *Identity in adolescence: Processes and contents*. San Francisco, CA: Jossey-Bass.

Wehrein, E., Lujan, H., DiCarlo, S. (2007). Gender differences in learning style preferences among undergraduate physiology students. *Advances in Physiology Education, 31*, 153-157.

Wolniak, G. (2013). Examining STEM bachelor's degree completion for students with differing propensities at college entry. *Journal of College Student Retention: Research, Theory, & Practice, 5,* 1-23.

Xie, Y. & Shauman, K.A. (2003). *Women in science: Career processes and outcomes.* MA: Harvard University Press.

Zirkel, S. (2002). Is there a place for me? Role models and academic identity among white students and students of color. *Teachers College Record, 104*, 357-376.

Tables

Table 1.

Descriptive statistics (student-class observations)

	Full sample			Female			Male		
	Mean	SD	Sample Size	Mean	SD	Sample Size	Mean	SD	Sample Size
Student Demographic Characteristics									
Variable									
Ethnicity			20,264			10,734			9,476
White	0.13	0.33		0.12	0.32		0.14	0.35	
Black	0.01	0.12		0.02	0.02		0.01	0.11	
Hispanic	0.20	0.40		0.21	0.40		0.19	0.39	
Asian	0.50	0.50		0.52	0.50		0.48	0.50	
Other race	0.06	0.24		0.06	0.24		0.06	0.24	
Non-resident	0.09	0.29		0.08	0.26		0.11	0.31	
Low-income status	0.38	0.48	20,252	0.41	0.49	10,730	0.35	0.48	9,468
First-generation status	0.54	0.49	19,744	0.56	0.49	10,503	0.52	0.50	9,188
SAT math	626.51	83.70	19,385	608.84	82.09	10,427	647.09	80.82	8,940
SAT verbal	561.59	89.86	19,385	558.68	87.84		564.98	92.05	
HS GPA	3.91	0.26	20,244	3.93	0.24	10,725	3.89	0.27	9,465
Female teacher	0.44	0.49	20,292	0.49	0.50	10,734	0.38	0.49	9,476
Key Outcome Measures									
Grade in course	2.55	1.06	19,973	2.48	1.05	10,569	2.64	1.07	9,327
Passing grade (“C” or higher)	0.85		19,973	0.85		10,569	0.87		9,327
Outcome (instructor is female)									
Grade in course	2.52	1.08	8,830	2.47	1.06	5,225	2.57	1.11	3,570
Passing grade (“C” or higher)	0.84		8,866	0.84		5,244	0.84		3,586

	Full sample			Female			Male		
	Mean	SD	Sample Size	Mean	SD	Sample Size	Mean	SD	Sample Size
Outcome (instructor is male)									
Grade in course	2.58	1.05	11,143	2.48	1.04	5,344	2.67	1.04	5,757
Passing grade ("C" or higher)	0.88		5,783	0.86		5,370	0.88		5,783
Behavior engagement variables			11,274			6,770			4,395
Attended lectures	0	1		0.05	0.93		-0.08	1.09	
Listened attentively to lectures	0	1		-0.00	0.98		0.01	1.03	
Asked the professor or TA for help	0	1		-0.00	0.99		0.00	1.01	
Asked questions or contributed to the discussion in lecture	0	1		-0.06	0.98		0.08	1.02	
Utility value variables			11,274			6,770			4,395
Doing well in this course is important to me	0	1		0.05	0.97		-0.08	1.04	
Understanding the material in this course has many benefits for me	0	1		-0.04	1.01		0.07	0.98	
Efficacy variables			11,274			6,770			4,395
I'm certain I can master the skills taught in this class	0	1		-0.13	1.00		0.20	0.96	
I'm certain I can figure out how to do the most difficult course material	0	1		-0.14	1.00		0.21	0.96	
I can do almost all the work in the class if I don't give up	0	1		-0.05	1.02		0.08	0.96	
Interest variables			11,274			6,770			4,395
I find many topics in this course to be interesting	0	1		-0.06	1.01		0.09	0.97	
Solving problems in this class is interesting for me	0	1		-0.07	1.01		0.11	0.97	
I find this class intellectually stimulating	0	1		-0.03	1.01		0.05	0.98	

Table 2.

Summary statistics describing (1) the number of students enrolled in courses by department, student-class level and (2) the number of sections (n=73) taught by females within each department, instructor-class level

		Students		Instructors	
	Full sample Percent (n=20,292)	Female Percent (n=10,734)	Male Percent (n=9,476)	Female Percent (n=32)	Male Percent (n=42)
Biology	0.17	0.21	0.12	0.22	0.10
Chemistry	0.48	0.54	0.41	0.41	0.40
Engineering	0.06	0.03	0.11	0	0.17
Information & computer sciences	0.07	0.04	0.11	0	0.10
Math/statistics	0.16	0.14	0.18	0.28	0.17
Physics	0.05	0.02	0.07	0.06	0.07
Average number of courses taken	2.08	2.11	2.07		

Table 3.

Estimated role of female instructor status for student academic outcomes

	(1)	(2)	(3)	(4)
Panel A. Outcome measure: standardized course grade				
Instructor is female	-0.01	-0.07**	-0.05	---
	(0.03)	(0.03)	(0.04)	---
Student is female	-0.11***	-0.24***	---	---
	(0.02)	(0.05)	---	---
Interaction	0.04	0.01	0.01	0.04
	(0.04)	(0.03)	(0.04)	(0.05)
Panel B. Outcome measure: receiving a "C" or better	(1)	(2)	(3)	(4)
Instructor is female	-0.04*	-0.04***	-0.03**	---
	(0.02)	(0.01)	(0.02)	---
Student is female	-0.00	-0.01	---	---
	(0.02)	(0.03)	---	---
Interaction	0.02	0.02*	0.02	0.03*
	(0.01)	(0.01)	(0.01)	(0.02)
Course FE		X	X	
Student FE			X	X
Section FE				X
Student controls	X	X		

Notes. Each cell reports the coefficients from a separate linear probability regression. The standardized grade has a mean zero and standard deviation one. Models 1 through 4 include a control for instructor academic rank. Models 1 through 3 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 19,895. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Table 4.
Estimated role of female instructor status for student academic outcomes, high SAT-scoring students only

Panel A. Outcome measure: standardized course grade				
	(1)	(2)	(3)	(4)
Instructor is female	-0.04	-0.08**	-0.06	---
	(0.04)	(0.03)	(0.047)	---
Student is female	0.02	-0.20**	---	---
	(0.03)	(0.07)	---	---
Interaction	-0.01	0.06	0.04	0.06
	(0.05)	(0.06)	(0.05)	(0.06)
Panel B. Outcome measure: receiving a "C" or better				
	(1)	(2)	(3)	(4)
Instructor is female	-0.023	-0.041***	-0.03*	---
	(0.018)	(0.013)	(0.02)	---
Student is female	-0.003	-0.015	---	---
	(0.012)	(0.036)	---	---
Interaction	0.016	0.050**	0.04	0.04*
	(0.018)	(0.020)	(0.02)	(0.02)
Course FE		X	X	
Student FE			X	X
Section FE				X
Student controls	X	X		

Notes. Each cell reports the coefficients from a separate linear probability regression. The standardized grade has a mean zero and standard deviation one. Models 1 through 4 include a control for instructor academic rank. Models 1 through 3 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 6,961. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Table 5.
Estimates of the effect of instructor gender on non-academic outcomes: Behavioral engagement

Panel A. Attended lectures

	(1)	(2)
Instructor is female	0.23***	---
	(0.06)	---
Student is female	0.37***	0.37***
	(0.04)	(0.05)
Interaction	-0.05*	-0.08**
	(0.03)	(0.04)

Panel B. Listened attentively to lectures

	(1)	(2)
Instructor is female	0.09*	---
	(0.06)	---
Student is female	-0.03	-0.01
	(0.13)	(0.12)
Interaction	0.02	0.01
	(0.04)	(0.04)

Panel C. Asked the professor or TA for help

	(1)	(2)
Instructor is female	-0.10**	---
	(0.05)	---
Student is female	-0.22	-0.21*
	(0.12)	(0.11)
Interaction	0.19***	0.19***
	(0.05)	(0.04)

Panel D. Asked questions or contributed to the discussion in lecture

	(1)	(2)
Instructor is female	-0.08**	---
	(0.03)	---
Student is female	-0.11	-0.10
	(0.13)	(0.12)
Interaction	0.06*	0.07*
	(0.04)	(0.04)

Table 5, continued
Panel E. Behavioral engagement factor

Instructor is female	-0.00	---
	(0.04)	---
Student is female	-0.04	-0.03
	(0.16)	(0.15)

Interaction	0.11***	0.10***
	(0.04)	(0.04)
Course FE	X	
Section FE		X
Student controls	X	X

Notes. Each cell reports the coefficients from a separate linear probability regression. Models 1 through 3 include a control for instructor academic rank. Models 1 and 2 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 11,165. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Table 6.

Estimates of the effect of instructor gender on non-academic outcomes: Student interest

Panel A. I find many topics in this course to be interesting		
	(1)	(2)
Instructor is female	-0.10**	---
	(0.04)	---
Student is female	0.05	-0.12
	(0.10)	(0.11)
Interaction	0.08**	0.08**
	(0.04)	(0.04)
Panel B. Solving problems in this class is interesting for me		
	(1)	(2)
Instructor is female	-0.07**	---
	(0.03)	---
Student is female	-0.12*	-0.36***
	(0.06)	(0.04)
Interaction	0.03	0.05
	(0.03)	(0.03)
Panel C. I find this class intellectually stimulating		
	(1)	(2)
Instructor is female	-0.09*	---
	(0.05)	---
Student is female	-0.01	-0.13
	(0.07)	(0.09)
Interaction	0.06	0.07
	(0.04)	(0.05)
Panel D. Student interest factor		
	(1)	(2)
Instructor is female	-0.09*	---
	(0.05)	---
Student is female	-0.78***	-0.31***
	(0.09)	(0.06)
Interaction	0.02	0.01
	(0.04)	(0.05)
Course FE	X	
Section FE		X
Student controls	X	X

Notes. Each cell reports the coefficients from a separate linear probability regression. Models 1 through 3 include a control for instructor academic rank. Models 1 and 2 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 9,733. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Table 7.

Estimates of the effect of instructor gender on non-academic outcomes: Self-efficacy

Panel A: I'm certain I can master the skills taught in this class		
	(1)	(2)
Instructor is female	-0.03	---
	(0.03)	---
Student is female	-0.18*	-0.20***
	(0.08)	(0.02)
Interaction	-0.01	0.05
	(0.04)	(0.03)
Panel B: I'm certain I can figure out how to do the most difficult course material		
	(1)	(2)
Instructor is female	-0.05	---
	(0.04)	---
Student is female	-0.23***	-0.20***
	(0.06)	(0.02)
Interaction	-0.00	0.06*
	(0.05)	(0.04)
Panel C: I can do almost all the work in the class if I don't give up		
	(1)	(2)
Instructor is female	-0.02	---
	(0.04)	---
Student is female	0.11**	-0.06**
	(0.04)	(0.03)
Interaction	-0.06	-0.02
	(0.04)	(0.04)
Panel D: Self-efficacy factor		
	(1)	(2)
Instructor is female	-0.07*	---
	(0.04)	---
Student is female	-0.32***	-0.29***
	(0.05)	(0.03)
Interaction	-0.04	0.02
	(0.04)	(0.04)
Course FE	X	
Section FE		X
Student controls	X	X

Notes. Each cell reports the coefficients from a separate linear probability regression. Models 1 through 3 include a control for instructor academic rank. Models 1 and 2 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 9,749. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Appendix

Table A1.
Course-level descriptive statistics

Courses	Number of observations	Average section size	Number of sections	Number of female instructors	Average grade	
DNA to Organisms	1894	315.67	6	3	2.56	(1.02)
From Organisms to Ecosystems	1460	292	5	4	2.76	(0.89)
General Chemistry A	2440	406.67	6	3	2.45	(1.09)
General Chemistry B	1750	350	5	0	2.61	(0.93)
General Chemistry C	1369	342.25	4	2	2.50	(1.05)
Organic Chemistry A	1434	286.80	5	3	2.52	(1.08)
Organic Chemistry B	1201	300.25	4	2	2.51	(1.06)
Organic Chemistry C	1433	358.25	4	3	2.54	(1.00)
Computational Methods in ECE	201	201	1	0	2.93	(0.93)
Introductory Programming (EECS)	216	216	1	0	2.83	(1.01)
Introductory Engineering*	243	243	1	0	2.47	(1.01)
Statics*	370	370	1	0	2.67	(1.00)
Introductory Thermodynamics*	243	243	1	0	2.30	(0.81)
Introduction to Programming	888	296	3	0	3.18	(0.78)
Programming with Software Libraries*	449	449	1	0	2.67	(1.12)
Pre-Calculus*	214	214	1	1	2.19	(1.32)
Single-Variable Calculus 1	1228	204.67	6	5	2.14	(1.21)
Single-Variable Calculus 2	660	330	2	1	2.15	(1.17)
Multivariable Calculus	461	230.50	2	0	2.25	(1.27)
Introduction to Linear Algebra	234	234	1	0	2.31	(1.04)
Classical Physics	950	237.50	4	2	2.76	(0.97)
Psychology Fundamentals	263	263	1	1	2.41	(0.99)
Basic Statistics	371	185.50	2	2	2.93	(0.97)

Note. Standard deviation in parentheses. Courses with asterisks did not administer experience surveys.

Table A2.
Estimated effect of female instructor assignment on completing a survey

	(1)	(2)	(3)	(4)
Instructor is female	0.12	0.02	0.03	---
	(0.08)	(0.06)	(0.05)	---
Interaction	-0.07	0.02	-0.01	-0.01
	(0.04)	(0.03)	(0.03)	(0.02)
Observations	20209	20209	20209	20209
Course FE		X	X	
Student FE			X	X
Section FE				X
Student controls	X	X		

Notes. Each cell reports the coefficients from a separate linear probability regression. Models 1 through 4 include a control for instructor academic rank. Models 1 through 3 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 20,209. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.

Table A3.
Regressions of faculty gender on student characteristics

	All students		Female students	
	Dependent variable: student is female		SAT math score	HSGPA
	(1)	(2)	(3)	(4)
Instructor is female	-0.01	-0.01	4.73	0.02**
	(0.01)	(0.01)	(3.72)	(0.01)
Student controls		X		
N	20209	20209	10431	10725

Note. Standard errors in parentheses.

Table A4.
Estimated role of female instructor status for student academic outcomes (year 1 and 2 combined)

Panel A. Outcome measure: standardized course grade				
	(1)	(2)	(3)	(4)
Instructor is female	0.01	-0.01	0.02	---
	(0.02)	(0.02)	(0.03)	---
Student is female	-0.10***	-0.06	---	---
	(0.01)	(0.08)	---	---
Interaction	0.05*	0.01	0.02	0.04
	(0.03)	(0.02)	(0.02)	(0.02)
Panel B. Outcome measure: receiving a "C" or better				
	(1)	(2)	(3)	(4)
Instructor is female	-0.01	-0.02***	-0.01	---
	(0.01)	(0.01)	(0.01)	---
Student is female	-0.02***	0.02	---	---
	(0.01)	(0.02)	---	---
Interaction	0.02**	0.02*	0.02**	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)
Course FE		X	X	
Student FE			X	X
Section FE				X
Student controls	X	X		

Notes. Each cell reports the coefficients from a separate linear probability regression. The standardized grade has a mean zero and standard deviation one. Models 1 through 4 include a control for instructor academic rank. Models 1 through 3 include a term fixed effects. Student controls are ethnicity, first-generation status, low-income status, SAT verbal score, SAT math score, and high-school point average. Missing values have been adjusted using a dummy variable approach. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively. The sample size is 58,569. The interaction term represents a female-student-female-instructor interaction. Standard errors in parentheses and are clustered at the section level.