Economics of Education Review 61 (2017) 9–18

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Why does teacher gender matter?☆

Dario Sansone

Georgetown University, Department of Economics, ICC 580, 37th and O Streets, NW, Washington, DC 20057-1036, United States

ARTICLE INFO

Keywords:
High school
STEM
Teacher gender
Role model

JEL:
C33
I20
J16

ABSTRACT

This paper shows that high school math and science teacher gender affects student interest and self-efficacy in STEM. However, such effects become insignificant once teacher behaviors and attitudes are taken into account, thus pointing towards an omitted variables bias. Teacher beliefs about male and female ability in math and science – as well as how teachers treat boys and girls in the classroom – matter more than teacher's own gender. The student fixed effects estimates also highlight that creating a positive learning environment and making math and science interesting are pivotal in engaging students in these subjects.

1. Introduction

There is widespread concern among U.S. policy-makers regarding the gender gap in STEM (science, technology, engineering and math) fields. Indeed, female students perform consistently worse than their male peers in international evaluations such as the PISA (NCES, 2016), as well as in the SAT math test (College Board, 2016). Furthermore, they are less likely to major in these subjects (National Science Foundation, 2015). As a result, women are underrepresented among science and engineering professors (National Science Foundation, 2015), as well as in tech industries (Apple, 2015). These factors aggravate the overall shortage of workers in STEM fields (Carnevale, Smith, & Melton, 2011; Executive Office of the President, 2012).

As also summarized by Ceci, Ginther, Kahn, and Williams (2014), this gender gap has received considerable attention in recent years (Dennehy & Dasgupta, 2017; Fryer Jr & Levitt, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Mechtenberg, 2009; Schneeweis & Zweimüller, 2012). Within this literature, several scholars have linked student performances and career decisions with teacher gender both in primary and secondary schools (Antecol, Ozkan, & Serkan, 2015; Dee, 2007; Holmlund & Sund, 2008; Muralidharan & Sheth, 2016; Paredes, 2014; Winters, Haight, Swaim, & Pickering, 2013), as well as universities (Bettinger & Long, 2005; Bottia, Stearns, Mickelson, Moller, & Valentino, 2015; Carrell, Page, & West, 2010; Griffith, 2010; Hoffmann & Oreopoulos, 2009; Price, 2010).

Following the theoretical framework developed by Paredes (2014), teacher gender may affect students in a variety of ways; by acting as role models, reinforcing stereotype threats, and through teacher biases that are correlated with gender. First, students may perform better when assigned to a same-sex teacher if they identify themselves with such a role model. In other words, female students are exposed to successful women in STEM when assigned to female teachers, and may therefore be inspired by them to go into these fields. Second, students may also react to teacher gender by internalizing an expected negative stereotype about their gender. The resulting anxiety may reduce their academic performance.

Third, teacher gender may affect teacher behavior. Female teachers could impact student performances because they may have higher math anxiety (especially in primary schools), which may negatively affect students (Antecol et al., 2015). Moreover, female teachers may structure their classroom, select topics and provide examples differently than their male colleagues.

Within this context, teachers may also have their own gender biases, which may affect how they treat students and evaluate them (Lavy, 2008), thus impacting student performance. This study investigates the importance of this channel by including in the estimated model how teachers compare men and women in math and science, as well as by whether teachers treat male and female students differently. The former is particularly relevant since gender-related math and science attitudes may be transmitted from teachers to students (Gunderson, Ramirez, Levine, & Beilock, 2012).

☆ I am grateful to Laurent Bouton, Mary Ann Bronson, Caitlin Brown, Garance Genicot, Arik Levinson, Givi Melkadze, Elena Arias Ortiz, Mariacristina Rossi, Allison Stashko, Francis Vella, the Editor Daniel Rees, two anonymous referees and participants to the Georgetown EGSO seminar for their helpful comments. I am also grateful to John Rust and Judith House for their technical support. The usual caveats apply.

E-mail address: ds1289@georgetown.edu.

http://dx.doi.org/10.1016/j.econedurev.2017.09.004
Received 8 May 2017; Received in revised form 15 September 2017; Accepted 16 September 2017
Available online 19 September 2017

Although most of the aforementioned studies find positive effects of female teachers on female student achievements, the overall mixed results indicate a multifaceted issue. Indeed, female teachers represent a highly heterogeneous group, so it is unsurprising that the empirical conclusions are not clear-cut.

This paper tests whether and why high-school teacher gender matters. It clarifies whether teacher gender has an intrinsic value per se, e.g. because of a role model effect or a lower stereotype threat, or if rather what really matters in raising students' interest and self-confidence is how teachers treats them and manages the classroom. The novelty of this study is that the empirical analysis directly includes not only teacher gender, but also several direct measures of teacher behaviors. In particular, it is possible to incorporate whether teachers listen and values students' ideas, whether they make their subjects interesting, and whether they have high expectations for all their students.

Using a between-subject student fixed effects model, i.e. by comparing math and science teachers for the same students, I show that teacher gender affects student interest and self-confidence. However, such effects become statistically insignificant once I control for teacher behaviors, attitudes and expectations. Teacher gender matters because teachers differ in their gender-related math and science attitudes, as well as in how they treat male and female students. From an econometric point of view, this means that omitted variables bias is a key issue in estimating the effect of teacher gender.

In addition to this, I also show that students are influenced by teachers who listen to their ideas, who make their subjects interesting, and who have high expectations for all of their students. Therefore, this paper not only tests whether the estimates of teacher gender are biased because of omitted variables, but it also investigates which factors affect student interest and self-confidence in STEM. Finally, in the last part of the paper, I explore whether it is possible to identify teachers with desirable attitudes and behaviors from their observable characteristics in the resume.

This paper also adds to the literature by focusing on student interest and self-confidence rather than by just looking at test scores. These outcome variables are particularly important since scholars have suggested that anxiety towards mathematics and lack of self-confidence may be behind girls' underperformances in STEM subjects (OECD, 2015). Higher self-confidence is essential to "think like scientists": that is, to take risks and to invest in a trial-and-error process in order to accumulate additional knowledge in math and science. In addition to this, self-confidence is pivotal in enhancing intrinsic motivation (Koch, Nafziger, & Nielsen, 2015). Indeed, greater self-confidence increases students' expected productivity, thus raising their motivation to study and leading them to exert more effort. To further motivate the importance of analyzing how to increase self-efficacy, it is also worth mentioning that Filippin and Paccagnella (2012) show in a theoretical model how small initial differences in self-confidence can lead to diverging patterns of human capital accumulation among students with the same initial ability.

Enjoying coursework is one of the most important determinants of college major choice (Zafar, 2013). Indeed, many women report not pursuing careers in STEM because they deem the subject uninteresting (Weinberger & Leggon, 2004). It is therefore essential to analyze if and how high school teachers can affect student enjoyment and interest in STEM. To summarize, spurring curiosity about these subjects and boosting students' confidence in their own abilities are necessary conditions to increase the number of individuals who take advance math and science classes in high school and who major in STEM fields once in college.

This study is also the first to utilize the High School Longitudinal Study of 2009 (HSLS:09) to answer this question. Contrary to previous studies, the results presented in this paper provide information on a recent cohort, thus offering a new perspective on Millennials and their educational choices.

2. Theoretical framework

This section expands the theoretical model in Paredes (2014) in order to provide a formal framework and to clarify how teacher gender may affect student interest and self-efficacy.

2.1. Teachers time allocation

Teachers are assumed to choose how they allocate their limited time among students in order to maximize their own utility. Formally, the teacher objective function is the following:

$$\max_{h_{gt},h_{bt}} \alpha_{ts}(g)\frac{N_{gt}}{N_t}U(h_{gt}) + \alpha_{ts}(b)\frac{N_{bt}}{N_t}U(h_{bt})$$

Each teacher t in subject s (math or science)[1] has to decide how to allocate a fixed amount of hours h_t between girls (h_{gt}) and boys (h_{bt}). Each teacher has N_t students: N_{gt} girls and N_{bt} boys. The teacher utility function $U(.)$ is increasing in h, thus the teacher obtains more utility by working more with students.

The key parameter in this context is $\alpha_{ts}(.)$: it measures the teacher preferences towards same-sex students. It is equal to one if teacher and student have the same sex, while it is less than one otherwise. In other words, female teachers gain more utility from allocating time to female students, thus $h_{gt} > h_{bt}$.[2] An unbiased teacher would allocate time equally between boys and girls.

The advantage of the dataset used in this paper is that it includes two indicators of teacher bias from two different sources. Indeed, 9th grade teachers in math and science are asked to compare men and women in math and science. Furthermore, students are asked whether their math and science teachers treat boys and girls differently. This implies that, unlike in Paredes (2014), it is possible to measure such teacher bias and, most importantly, to allow it to vary by subject. As shown in the empirical analysis, the inclusion of such indicators is pivotal in understanding the mechanism behind the effect of teacher gender. These controls emphasize that gender-related teacher attitudes and behaviors, here symbolized by α_{ts}, matter more than teacher gender *per se*.

2.2. Student outcomes

Female student i's outcomes (interest and self-efficacy in subject s) result from the following function:

$$y_{its} = h(f_g(h_{gt}, h_{bt}),\ \beta_i r_{it},\ z_{it}'\gamma_i, a_i) +\ \varepsilon$$

$$f_g(h_{gt}, h_{bt})\ =\ \frac{h_{gt}}{N_g}\ +\ \rho_g\frac{h_{bt}}{N_t}$$

So each student outcome depends by how the teacher allocates time between boys and girls. Since ρ_g is less than 1, female students benefit more from the time allocated to them (h_{gt}/N_g) than from the time allocated to boys ($\rho_g h_{bt}/N_t$). Teacher gender may also have a direct effect (r_{it}), with a positive impact if $\beta_i > 0$. Note that such coefficient may be positive because of a role model effect, as well as because of the absence of a stereotype threat. Within this model, it is not possible to distinguish among these two channels, although some suggesting evidence is provided in Section 5.3. In addition to this, student interest and self-

[1] There is a little abuse of notation in this model: contrary to Section 3, here the notation includes both teacher t and subject s. This has been done to stress the heterogeneities at the teacher level within each subject that are possible to capture in the empirical analysis.

[2] A special case is represented by gender-segregated schools. However, only 4% of the HSLS:09 sample belongs to such category, so it is not the focus of this discussion.

confidence are also affected by teacher behaviors[3] (z_{it}) and by the student own ability (a_i). Some random shocks are captured by ε. A similar function can be written for male students. This model expands Paredes (2014) by including teacher behaviors (whether she listens to students' ideas), quality (whether she makes her subject interesting), and expectations (whether she believes that all students can be successful). Therefore, the model allows teachers to affect student outcomes y_{its} not only because of some intrinsic value of their gender (r_{it}), or by deciding how to allocate their time (h_{gt}, h_{bt}), but also through their in-class behaviors (z_{it}).

3. Data

3.1. High school longitudinal study of 2009

The HSLS:09 is a nationally representative panel database including around 26,000 students from 940 private and public schools in 2009. The survey design has two levels: first, schools were randomly drawn at the national level. Second, around 30 students in each school were randomly selected among 9th graders.[4] In the first round (Fall 2009), information was collected from the students, their parents, school administrators and lead school counselors.[5] Students and parents were interviewed again in 11th grade (Spring 2012) and at the beginning of college (Summer 2013). The key feature of this database is that, for each 9th grader, both the math and the science teachers were interviewed in the baseline survey.[6] Furthermore, students were asked to evaluate both teachers. As explained in the empirical section, the identification strategy is based on this duality.

The data are very rich and include student demographics, expectations, behaviors, attitudes, evaluations of the math and science teachers, as well as family background. In addition to this, a math assessment was administered to the students in 9th grade and in 11th grade. Data are also available from student transcripts including their GPA, AP class grades, SAT scores, and the number of credits taken in each subject during high school. Additional documentation about the HSLS:09 can be found in the Online Appendix, as well as in Ingels et al. (2011, 2014, 2015).[7]

3.2. Variables description

In order to capture student interest and self-confidence in math and science, I use three different outcome variables. The first is an indicator as to whether the 9th grader enjoys her math/science class. The second alternative measure of student interest is whether the 9th grader's favorite subject is math/science. Furthermore, I consider a measure of math and science self-efficacy (standardized to a mean of 0 and standard deviation of 1) constructed by applying principal component analysis to multiple inputs: whether the 9th grader is confident that she can do an excellent job in the math/science tests and assignments, master the skills in these courses, as well as whether she is certain that she can understand the textbooks in these classes. In other words, self-efficacy quantifies the level of student confidence in her own ability to perform certain tasks.[8]

Table 1
Student characteristics by subject and gender.

Variable	Math			Science		
	Male	Female	Diff	Male	Female	Diff
Enjoy Subject	0.67	0.66	0.01	0.70	0.66	0.04***
Favorite Subject	0.15	0.16	0	0.10	0.09	0.01***
Self-efficacy	0.14	−0.05	0.19***	0.16	−0.08	0.24***
Listens student ideas	0.86	0.85	0.01	0.87	0.85	0.02**
Makes subject interesting	0.64	0.61	0.03***	0.73	0.68	0.05***
All students can succeed	0.92	0.93	−0.01	0.91	0.92	−0.01
Treats boys/girls differently	0.14	0.09	0.05***	0.15	0.10	0.05***
Observations	10,670	10,430		10,670	10,430	

Teacher ability, behaviors and expectations are measured using four variables. Students are asked whether their math and science teachers value and listen to their ideas, whether they treat male and female students differently, whether they make their subject interesting, and whether they think that every student can be successful.[9] Students are also reminded that the survey is anonymous and that their principal and teachers have no access to their answers, thus reducing any concern about measurement error. In addition to this, teachers are also directly asked to compare males and females in math and science.

Previous studies have highlighted the importance of these teacher characteristics. Indeed, Fryer (2014) and Papageorge, Gershenson, and Kang (2016) have stressed the impact of teacher expectations on student achievements, while the impact of positive learning environments have been analyzed in Church, Elliot, and Gable (2001) and Lizzio, Wilson, and Simons (2002). Lavy and Sand (2015), have instead pointed out the short and long term effects of teachers' gender biases.

3.3. Descriptive statistics

Table 1 shows relevant summary statistics by subject and gender for students. There are no substantial differences between male and female students when looking at math enjoyment and interest, while male students are more likely to enjoy and be interested in science. The most remarkable difference between boys and girls can be found by looking at their self-efficacy. As also highlighted in other surveys (OECD, 2015), female students have a much lower average self-efficacy than male students. Moreover, the standard deviations are similar for both males and females, thus even the females at the top of the distribution have a lower average self-efficacy rate.

Teacher ratings also differ by student gender: male students are more likely to report that their teachers in 9th grade listen to students' ideas and make their subject interesting. They also tend to indicate more frequently that their teachers do not treat boys and girls equally, while there is no significant difference in the evaluations of teacher expectations.

In a similar way, Table 2 shows additional summary statistics by subject and gender for teachers. First of all, it is important to stress that 61% of math teachers and 56% of science teachers in the sample are female. Female math teachers are more likely to have an advanced degree and to have a regular certificate to teach in high school, while

[3] It is important to note that these behaviors are choice variables. In other words, teachers can decide how much effort to put in listening to students or in making their subject interesting. Therefore, student evaluations of teacher behavior (which is the what is observed, as explained in the next section) depends on teacher effort, as well as teacher ability. The latter is taken into account in the empirical analysis by including teacher characteristics such as education and experience.

[4] The Online Appendix discusses how this survey design has been taken into account in the empirical analysis.

[5] The parent questionnaire was completed by the parent or guardian most familiar with the 9th grader's school situation and experience.

[6] If the 9th grader had more than one science or math teacher, one teacher per subject was randomly selected among those teachers.

[7] Since the data are not publicly available, all sample size numbers are rounded to the nearest 10 for security reason.

[8] The Online Appendix includes a detailed description of all the variables used in the analysis, as well as additional robustness checks. When not shown in tables, results are available upon request.

[9] It is worth mentioning that the question on gender discrimination is asked after those regarding whether the teacher listens to students or whether she thinks that everybody can succeed, so there is no risk of priming students to think primarily about gender when answering those questions.

Table 2
Teacher characteristics by subject and gender.

Variable	Math			Science		
	Male	Female	Diff	Male	Female	Diff
More than Bachelor	0.49	0.52	−0.03***	0.58	0.56	0.02***
STEM major	0.42	0.39	0.03***	0.55	0.60	−0.05***
Experience	10.96	9.92	1.04***	12.18	9.85	2.33***
HS Certified	0.77	0.8	−0.03***	0.82	0.79	0.03***
Education degree	0.57	0.68	−0.11***	0.55	0.57	−0.02*
Listens student ideas	0.87	0.86	0.01*	0.88	0.86	0.02***
Makes subject interesting	0.65	0.62	0.03***	0.73	0.69	0.04***
All students can succeed	0.93	0.93	0	0.92	0.92	0
Boys better in math/science	0.14	0.08	0.06***	0.11	0.07	0.04***
Treats boys/girls differently	0.12	0.11	0.01***	0.13	0.12	0.01
Observations	7010	10,860		7100	9160	

the opposite is true in science. Only 40% of the math teachers and 58% of the science teachers majored in a STEM field,[10] two rather sobering indicators. Within each subject, this major choice is more common among male math teachers and female science teachers. On average, male teachers are more experienced than their female colleagues in both subjects. Female teachers are more likely to have their highest degree in education.

The second part of Table 2 highlights that male and female teachers significantly differ in their attitudes and behaviors. Therefore, as investigated in the next section, neglecting to take them into account could lead to omitted variables bias when estimating the impact of teacher gender on student interest and self-efficacy. Male teachers are more likely to believe that men are better than women in math or science. Female teachers are reported more frequently to treat all students equally, while students indicate more often that male teachers listen to students' ideas and make their subject interesting. Overall, math teachers are less successful than science teachers in making their subject appealing to students. There are no differences across gender and subject in whether teachers have high expectations for all their students.[11]

4. Main results

Similarly to Dee (2005) and Gershenson, Holt, and Papageorge (2016), my identification strategy relies on the fact that the HSLS:09 includes information about the math (M) and science (N) teachers for each 9th grader. It is then possible to estimate the following student fixed effects specification for male and female students separately:

$$y_{is} = \beta tgender_{is} + x_{is}'\gamma_1 + w_{is}'\gamma_2 + z_{is}'\gamma_3 + \mu_i + \alpha_s + \varepsilon_{is} \ \forall \ s \in \{M, N\}$$

The left-hand side variable (y_{is}) is the measure of student *i*'s interest or self-efficacy in subject *s*. The error term is defined as ε_{is}. The key regressor of interests highlighted in the literature is the math/science teacher gender ($tgender_{is}$). As usual, this equation also includes teacher education and experience (x_{is}). Since each student is observed twice – in math and science – it is also possible to include observable and unobservable student fixed effects (μ_i), as well as subject fixed effects (α_s). The main contribution of this paper is to investigate the effect of including additional variables omitted in previous studies. In particular, the HSLS:09 provides information on teacher ability, expectations and behaviors (z_{is}). Gender attitudes and teacher behaviors towards male and female students (w_{is}) are also recorded.[12]

By taking the difference between these two equations for math and science, it is possible to control for observable and unobservable variables that are constant across subjects at the individual level. This includes not only student individual characteristics such as race, cognitive/non-cognitive skills, and bias towards male or female teachers, but also school characteristics and family background. Nevertheless, it is important to note that the individual fixed effects do not account for subject-specific individual ability. For this reason, the estimated models include also whether the student earned an A in her math/science classes in middle school.[13]

The estimated coefficients are reported in Tables 3 and 4 for female and male students respectively. The dependent variable in Columns 1–3 is whether the student enjoys her math/science course in 9th grade. Whether the 9th grader's favorite subject is math or science is the dependent variable in Columns 4–6, while Columns 7–9 show the results for student self-efficacy.

Previous studies have not included teacher behaviors and attitudes (z_{is} and w_{is}). The omission of such variables leads to conclude that female teachers boost confidence among girls (Table 3 Column 7), while they reduce interest in STEM among boys (Table 4 Column 1 and 4). In other words, these between-subject student fixed effects models replicate the finding in most of the previous literature: teacher gender matters. In term of magnitudes, female teachers raise self-confidence in female students by at most 0.05 standard deviation on average, which is one fifth of the gender gap in self-efficacy. For comparison, Paredes (2014) finds that female teachers increase female students' performances by 0.04 standard deviations on average, which is one fourth of the gender gap in math.

The remaining specifications show which variables drive these results. Teacher ability, expectations and behaviors (z_{is}) are included in Columns 2, 5 and 8. Adding these regressors does not change the impact of teacher gender on self-efficacy for female students (Table 3 Column 7) and on enjoyment for male students (Table 4 Column 2). On the other hand, the effect of teacher gender is no longer statistically significant when looking at boys' favorite subject (Table 4 Column 5).

Gender attitudes and teacher behaviors towards male and female students (w_{is}) are included in Columns 3, 6 and 9. Once these variables are taken into account, the effect of teacher gender becomes insignificant in all specifications, both for boys and girls. Female *and* male students show less interest when their teachers treat them differently based on gender.[14] It is then possible to conclude that teacher beliefs

[10] By far, the most common STEM majors are Mathematics and Statistics among math teachers, Biology and Physics among science teachers.

[11] In line with the literature on biases against female teachers in student evaluations (Boring, 2017), female teachers are evaluated more negatively by both male and female students. For instance, boys are more likely to report that the teacher listens to students' ideas and makes the subject interesting when such teacher is a man. For later reference, it is important to note that this is true both in math and science. Similar differences in evaluations can also be found among female students. However, there are no statistical differences in male student evaluations when they are asked whether their math or science teachers treat boys and girls differently.

[12] The Online Appendix includes a detailed description of all the variables used in the analysis.

[13] In addition to this, it is possible to argue that, while these subject specific components could be important when comparing hard science with humanities (as in the previous literature), it does not seem that between math and science there is a substantial difference. For instance, (Patterson & Kobrin, 2012) reports a high correlation between the SAT scores in Math and Chemistry (0.756) or Physics (0.755). As an additional robustness check, I have also estimated the same models as in Tables 3 and 4, but excluding students who are taking biology classes, since such classes are typically less math-intense, thus the individual fixed effects may not capture student ability. Results are in line with the previous findings.

[14] These results are robust to different sample sizes. Indeed, the conclusions do not change if all specifications with the same outcome variable are estimated using the same sample size. The only exception is for female self-efficacy. The coefficients of female teacher in Columns 7 and 8 are still significant if I impose the same sample size of a specification which includes also whether teachers treat boys/girls differently. However, they become slightly insignificant if I impose the same sample size of the specification which includes also how teachers compare boys and girls in math/science (as in Column 9).

Table 3
Effect on female students of teacher gender.

	Enjoy Subject			Favorite Subject			Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female teacher	-0.015	-0.009	-0.015	0.005	0.004	0.006	0.050*	0.065**	0.042
	(0.017)	(0.013)	(0.014)	(0.011)	(0.012)	(0.013)	(0.028)	(0.026)	(0.029)
Listens student ideas		0.128***	0.133***		0.012	0.007		0.131***	0.152***
		(0.021)	(0.022)		(0.015)	(0.016)		(0.041)	(0.045)
Makes subject interesting		0.397***	0.385***		0.132***	0.132***		0.413***	0.427***
		(0.015)	(0.017)		(0.012)	(0.013)		(0.028)	(0.031)
All students can succeed		0.116***	0.101***		0.037*	0.043*		0.204***	0.207***
		(0.025)	(0.028)		(0.019)	(0.022)		(0.050)	(0.056)
Boys better in math/science			0.015			-0.009			-0.050
			(0.019)			(0.020)			(0.042)
Treats boys/girls differently			-0.053**			-0.043**			0.043
			(0.027)			(0.021)			(0.051)
Subject math fixed effect	-0.005	0.022**	0.021**	0.055***	0.060***	0.064***	0.064***	0.087***	0.093***
	(0.012)	(0.009)	(0.011)	(0.007)	(0.008)	(0.009)	(0.020)	(0.020)	(0.022)
A in 8th grade math/science	0.100***	0.082***	0.086***	0.115***	0.123***	0.139***	0.414***	0.392***	0.417***
	(0.018)	(0.015)	(0.016)	(0.013)	(0.014)	(0.016)	(0.034)	(0.032)	(0.035)
Constant	0.665***	0.165***	0.192***	0.056***	-0.095***	-0.110***	-0.272***	-0.863***	-0.892***
	(0.024)	(0.030)	(0.034)	(0.015)	(0.026)	(0.029)	(0.042)	(0.064)	(0.071)
Observations	13,270	13,050	11,640	14,530	12,970	11,560	13,080	12,880	11,490
Overall R^2	0.02	0.26	0.26	0.02	0.04	0.04	0.10	0.17	0.18
Within R^2	0.01	0.25	0.24	0.03	0.06	0.07	0.04	0.13	0.13

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
SE in parentheses clustered at school level. Source: HSLS09. Only female students considered. Additional controls not reported: teacher has more than Bachelor's degree, teacher has Bachelor's degree with STEM major, teacher's experience teaching math/science (squared).

about female abilities in math and science - together with teacher discriminatory behavior - are more important than teacher own gender in raising students' interest and confident in STEM fields.[15]

These results are supported from Table 5. While female teachers do not have a significant impact on average, having a good (or bad) teacher who is a woman could actually make a difference. This statement is tested in Table 5, which includes the interactions between teacher gender and all the variables describing teacher ability, behaviors and expectation. Almost all interaction terms have a coefficient statistically indistinguishable from zero. Nevertheless, one of the few exceptions holds for the interaction with teacher beliefs: students have lower self-efficacy when their female teachers believe that males are better than females in math or science. This is found for students of both sexes. What a male teacher believes does not matter. This implies that female teachers can actually have *larger* negative effects than their male colleagues if they perpetuate gender stereotypes. The other exception is also related to gender issues: if a female teacher treats boys or female differently, this negatively affects the course enjoyment among her male students. To summarize, these statistically significant interaction terms emphasizes that teachers' prejudices and discriminatory behaviors harm both male and female students by affecting their confidence and enjoyment in STEM courses.

From a policy perspective, it is important to note that all students are positively affected when teachers value and listen to their ideas, as well as when teachers make their subjects interesting. It is also worth mentioning that girls show more interest and higher self-efficacy when their teachers believe that all students could be successful. All of these results are particularly relevant when designing teacher hiring and training policies.

5. Additional results

5.1. Heterogeneity

I also investigate whether the effects of teacher gender and behavior change by subject, thus estimating β_s and γ_s instead of just β and γ. There is no difference in the effect of teacher gender between math and science. On the other hand, whether the teacher listens to students' ideas raises the probability that a male student reports math as his favorite subject, while it increases self-efficacy for girls more in math than science. Whether the teacher makes the subject interesting increases the probability that students enjoy the course more in science (both for boys and girls) and makes female students more confident in science, while it raises the probability that a male student reports math as his favorite subject.

There could also be heterogeneous effects by ability: teachers could have a different impact on top students and on low-achievers. To identify top students, I have used the scores from the math test administered to 9th graders during the survey. Therefore, I have estimated the same specifications as the ones in Tables 3 and 4, but only for students whose standardized theta score is above the overall median. The results for the high-achievers are quantitatively similar to those for the whole sample.

Similar results are obtained by interacting teacher gender and behavior with whether the student got an A in her 8th grade math/science class. Almost all interaction terms are not statistically significant. One exception when looking at student favorite subject is the interaction with whether the teacher makes the subject interesting, which has a bigger effect on these top students.

Finally, it is important to stress that there is a reason behind the decision to report the results for two variables measuring student interest in math and science (student course enjoyment and favorite

[15] As shown in the descriptive statistics, female and male students evaluate teachers differently. However, this is not an issue since the results are presented for boys and girls separately. As already discussed, teacher gender may also affect student evaluation. In other words, teacher gender and behaviors may be related not only because male and female teachers behave differently, but also because they are evaluated differently by their student. Nevertheless, there is no evidence that such biases differ across subjects both in our sample and in other studies (Boring, 2017), so they are captured by the between-subject student fixed effects. Furthermore, even if such biases were not taken into account, the main result would not change: whether it is because of actual or perceived differences in behaviors, teacher gender is statistically significant only when such variables are omitted. In addition to this, controlling for these evaluations implies, given the bias, that we are comparing extremely fair female teachers with fair male teachers (since the other female teachers would be evaluated as unfair): therefore, if this indirect channel were driving the results, we would actually expect an even bigger effect of teacher gender, which is the opposite of what we find. Finally, the coefficient of teacher gender remains statistically insignificant even when we omit this control but we include all the other measures of teacher behavior.

Table 4
Effect on male students of teacher gender.

	Enjoy Subject			Favorite Subject			Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female teacher	-0.053***	-0.027**	-0.021	-0.021*	-0.015	-0.019	-0.001	0.021	0.018
	(0.015)	(0.012)	(0.013)	(0.011)	(0.011)	(0.012)	(0.024)	(0.024)	(0.025)
Listens student ideas		0.177***	0.189***		0.008	0.017		0.124***	0.124***
		(0.021)	(0.024)		(0.015)	(0.016)		(0.041)	(0.046)
Makes subject interesting		0.384***	0.379***		0.134***	0.138***		0.367***	0.380***
		(0.016)	(0.017)		(0.012)	(0.013)		(0.028)	(0.030)
All students can succeed		0.025	0.021		-0.006	-0.016		0.097*	0.068
		(0.026)	(0.029)		(0.020)	(0.023)		(0.057)	(0.063)
Boys better in math/science			0.015			-0.019			0.015
			(0.021)			(0.021)			(0.044)
Treats boys/girls differently			-0.061**			0.028			0.025
			(0.025)			(0.019)			(0.045)
Subject math fixed effect	-0.020*	0.012	0.004	0.042***	0.053***	0.051***	0.013	0.041**	0.036**
	(0.011)	(0.009)	(0.009)	(0.007)	(0.008)	(0.009)	(0.017)	(0.016)	(0.018)
A in 8th grade math/science	0.098***	0.086***	0.083***	0.108***	0.103***	0.082***	0.333***	0.320***	0.307***
	(0.015)	(0.014)	(0.015)	(0.012)	(0.014)	(0.015)	(0.031)	(0.031)	(0.033)
Constant	0.707***	0.233***	0.226***	0.097***	0.004	0.010	0.074**	-0.404***	-0.414***
	(0.021)	(0.032)	(0.037)	(0.015)	(0.023)	(0.026)	(0.035)	(0.066)	(0.075)
Observations	13,190	12,940	11,520	14,600	12,810	11,410	12,960	12,750	11,350
Overall R^2	0.02	0.25	0.26	0.02	0.04	0.04	0.10	0.15	0.15
Within R^2	0.01	0.23	0.24	0.02	0.05	0.04	0.03	0.09	0.09

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
SE in parentheses clustered at school level. Source: HSLS09. Only male students considered. Same controls as Table 3.

subject). The first, whether the student enjoys the course, is a "flow" variable: it refers to a specific class, thus the impact of a single teacher should be larger. On the other hand, the second, whether the student favorite subject is math or science, represents a "stock" variable: those preferences have been built over the years, thus one semester is less likely to change them, unless the student has an exceptional (good or bad) teacher. These observations are consistent with the coefficients reported in Tables 3 and 4. For instance, whether the teacher makes the subject interesting has a larger impact on student enjoyment (Columns 3 versus Columns 6), both for boys and girls.[16]

5.2. Can we identify good teachers from their resume?

Looking at Tables 3 and 4, it is possible to conclude that the variables that positively affects both male and female students' interest and confidence in almost all specifications are whether the teacher values and listens to students' ideas, as well as whether the teacher makes the subject interesting. The next step is to test whether it is possible to identify these high performing teachers through their observable features. This is done in Tables 6–10, which list the average characteristics of the high performing teachers together with those of the low performing teachers, for math and science separately. For instance, Table 6 compares math (science) teachers who think that boys are better than girls in math (science) with those who think the opposite. It also tests whether such differences are statistically significant. Tables 7–10 repeat the same procedure for all the other key regressors from Table 3.

As shown in Table 6 and also discussed in Section 2, female teachers are substantially less likely to think that boys are better than girls in math or science. The same can be said about teachers with a STEM major, with more experience, and certified to teach in high school. Math teachers who have more than a bachelor degree are also less likely to think that boys are better, while the opposite is true with science teachers whose highest degree is in Education.

As far as the other indicators are concerned, although most of the differences are statistically significant, the magnitude is rather small. For instance, the average years of experience teaching math or science in high school is lower among teachers who listen to their students' ideas, do not treat boys and girls differently, make their subject interesting, and think that all their students can succeed. However, such differences are always smaller than 1 year. In line with the findings in Harris and Sass (2011), it is remarkable that teachers with a graduate degree do not behave differently than those with only a Bachelor's degree. Consistently with the previous literature (Kane, Rockoff, & Staiger, 2008), there are also no striking differences among teachers based on their certification status. Certified teachers are slightly less likely (4 percentage points) to treat students differently based on their gender, but they also tend to have lower expectations for their students. To summarize, these formal measures of quality are not enough to signal top teachers in this context.[17] Another way to present this finding is to estimate a Probit model for each teacher behavior in Tables 6–10 on teacher gender and observable characteristics. Except when the dependent variable is how teachers compare males and females in math and science, all the other specifications have an extremely low McFadden-R^2 (less than 0.5%).

Consistently with the above conclusions, even if the main specifications already include controls for whether the teacher has more than a Bachelor's degree, for whether her major in college is in a STEM field, and for her experience teaching the subject in high school,[18] all these variables do not significantly affect their students' interest and self-efficacy. For this reason, I have also tried to add as controls whether the

[16] A more comprehensive measure of the student interest in these subjects can be obtained by combining through principal component analysis several answers (including the two used in the main tables): not only whether the student enjoys those classes or whether math/science is her favorite subject, but also if she thinks that those courses are a waste of time or boring, whether her least favorite subject is math/science, and whether she is taking those classes because she enjoys math/science. Consistently with the previous results, without controlling for teacher behaviors, expectations and attitudes, female teacher reduced male student interest by .1 standard deviations. Once these additional controls are included, teacher gender is no longer significant: the pivotal factors are whether the teacher makes the subject interesting, as well as whether she listens to students.

[17] Similar conclusions can be reached by considering only the evaluations from female students. The main difference is that there is a larger gap in term of experience (1-2 years) when looking at whether teachers have desirable attitudes, behaviors, and expectations. Female students are also more likely to report that their female teachers treat all students equally (up to 9 percentage points difference).

[18] Following the approach stressed in the educational literature, nonlinearities are taken into consideration by including experience as a polynomial of grade two.

Table 5
Interactions with female teacher.

	Female students			Male students		
	(1) Enjoy	(2) FavSubj	(3) Efficacy	(4) Enjoy	(5) FavSubj	(6) Efficacy
Female teacher	-0.005	0.085**	0.114	-0.024	-0.015	0.203*
	(0.058)	(0.040)	(0.111)	(0.055)	(0.037)	(0.117)
Listens student ideas	0.159***	0.036	0.210***	0.182***	-0.002	0.085
	(0.034)	(0.025)	(0.068)	(0.036)	(0.024)	(0.072)
& Interact with female teacher	-0.043	-0.047	-0.100	0.011	0.034	0.064
	(0.045)	(0.032)	(0.086)	(0.045)	(0.033)	(0.090)
Makes subject interesting	0.407***	0.119***	0.376***	0.401***	0.159***	0.436***
	(0.024)	(0.020)	(0.043)	(0.028)	(0.021)	(0.047)
& Interact with female teacher	-0.035	0.021	0.083	-0.038	-0.037	-0.095
	(0.032)	(0.025)	(0.057)	(0.034)	(0.027)	(0.058)
All students can succeed	0.076*	0.073**	0.224***	-0.005	-0.008	0.172**
	(0.042)	(0.033)	(0.071)	(0.041)	(0.034)	(0.085)
& Interact with female teacher	0.046	-0.053	-0.019	0.042	-0.011	-0.178
	(0.053)	(0.041)	(0.102)	(0.053)	(0.040)	(0.114)
Boys better in math/science	0.003	-0.003	0.051	0.044	-0.007	0.083
	(0.031)	(0.027)	(0.063)	(0.031)	(0.030)	(0.054)
& Interact with female teacher	0.025	-0.012	-0.204**	-0.057	-0.028	-0.147*
	(0.048)	(0.038)	(0.103)	(0.042)	(0.041)	(0.081)
Treats boys/girls differently	-0.084**	-0.034	0.039	-0.010	0.007	0.010
	(0.038)	(0.029)	(0.079)	(0.032)	(0.026)	(0.064)
& Interact with female teacher	0.063	-0.015	0.013	-0.087**	0.035	0.024
	(0.051)	(0.041)	(0.102)	(0.037)	(0.033)	(0.077)
Subject math fixed effect	0.022**	0.064***	0.091***	0.003	0.050***	0.034*
	(0.011)	(0.009)	(0.021)	(0.009)	(0.009)	(0.018)
A in 8th grade math/science	0.087***	0.139***	0.419***	0.082***	0.081***	0.305***
	(0.016)	(0.016)	(0.034)	(0.015)	(0.015)	(0.033)
Constant	0.182***	-0.156***	-0.937***	0.229***	0.007	-0.521***
	(0.049)	(0.037)	(0.091)	(0.048)	(0.033)	(0.099)
Observations	11,640	11,560	11,490	11,520	11,410	11,350
Overall R^2	0.26	0.04	0.18	0.25	0.04	0.15
Within R^2	0.25	0.07	0.14	0.24	0.05	0.09

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
SE in parentheses clustered at school level. Source: HSLS09. Same controls as Table 3. The interaction terms are between the female teacher indicator and the regressor above each interaction line.

teacher has a regular certification to teach in high school. The estimated coefficient is not statistically different from zero and the main results regarding gender and behaviors do not change. The same conclusions can be reached by adding an indicator variable equal to one if the teacher has her highest degree in Education.[19]

From a policy perspective, it would also be helpful if it were possible to identify specific pedagogical tools that would lead to higher interest and self-efficacy among students. In order to achieve this goal, I have estimated the same specifications as in Tables 3 and 4 while adding whether the teacher has students working in small groups. However, the main conclusions do not change and this variable is never significant in all regressions, even when looking at how groups are formed (e.g. by creating groups with students of similar ability levels).

5.3. A deeper look into teacher gender

The aim of this section is to provide further evidence on the role of the teacher gender. The specifications presented in Tables 3 and 4 (controlling for teacher behaviors, beliefs and expectations) can also be estimated by OLS, that is, separately for each subject, thus without student fixed effects. Also in these specifications, these simple correlation coefficients are small and insignificant when looking at teacher gender. Such coefficients are statistically significant only when analyzing female student self-efficacy. In this case, the magnitude is around 0.07 standard deviations.

Students are also asked whether they have talked with a teacher about which math or science courses to take during their first year of high school. If female teachers were stronger role models than males,

[19] On the other hand, higher teacher self-efficacy (which measures whether the teacher thinks that student performances are not mainly due to family background, as well as whether the teacher thinks that she has the ability to tackle difficult situations) leads to higher enjoyment and interest in the subject among girls, but it does not affect neither male students nor self-efficacy among female students.

Table 6
Teacher characteristics (Mean) – Boys better in math/science.

Variable	Math			Science		
	No	Yes	Diff	No	Yes	Diff
Female	0.63	0.48	0.15***	0.57	0.46	0.10***
More than Bachelor	0.52	0.44	0.08***	0.57	0.58	−0.01
STEM major	0.41	0.38	0.03**	0.59	0.47	0.12***
Experience	10.43	8.93	1.50***	10.99	10.02	0.97***
HS Certified	0.79	0.76	0.03***	0.82	0.75	0.07***
Education degree	0.64	0.66	−0.02	0.56	0.65	−0.08***
Observations	14,310	1690		13,210	1310	

Table 7
Teacher characteristics (Mean) – Listens student ideas.

Variable	Math			Science		
	No	Yes	Diff	No	Yes	Diff
Female	0.62	0.6	0.02*	0.6	0.56	0.04***
More than Bachelor	0.51	0.51	0	0.58	0.57	0.01
STEM major	0.42	0.4	0.02	0.55	0.59	−0.04***
Experience	11.1	10.32	0.78***	11.57	10.84	0.73***
HS Certified	0.81	0.78	0.03***	0.82	0.8	0.01
Education degree	0.51	0.54	−0.04***	0.56	0.56	0
Observations	2030	12,450		1690	11,060	

we would expect female students to talk more with female teachers about these educational choices. Nevertheless, the percentage of students reporting to discuss course selection with a teacher is the same for female students with a male or female math teacher. Percentages are similar also for science teachers. This supports the conclusion from Section 3 that teacher gender does not have an impact on students because of gender *per se*.

Similarly, as already discussed, one may argue that whether the student reports that the teacher makes the subject interesting may depend on the teacher gender: female teachers may adjust the content of their courses to include topics and examples which may raise the curiosity of female students. For instance, science could be more attractive to girls if they understood the impact that they would have on society. Science teachers are indeed asked how much emphasis they are placing on teaching students about the relationship between science, technology and society. In contrast to the above reasoning, female teachers report more frequently than their male colleagues to put minimal or no emphasis on such goal.

In conclusion, the results from this section supported the main findings: teacher gender has only a minimal (possibly null) direct effect on student interest and confidence. Creating a positive learning environment, treating all students equally, and transmitting passion about the subjects are the key drivers. One may argue that high school is too late, that female teachers can have a much larger effect in elementary schools, when gender gaps seem to emerge (Fryer Jr & Levitt, 2010), or that mentors are key mainly in tertiary education (Pollack, 2013), where women are underrepresented among faculty members. However, most of the literature highlighted in the introduction focuses on secondary education. Therefore, in response to such strong policy advice, this paper offers a cautionary note on putting too much hope on the effect of hiring more high school female teachers if the quality, efforts and beliefs are left unchanged.

Table 8
Teacher characteristics (Mean) – Treat boys/girls differently.

Variable	Math			Science		
	No	Yes	Diff	No	Yes	Diff
Female	0.61	0.57	0.04***	0.57	0.56	0.01
More than Bachelor	0.51	0.49	0.02	0.57	0.55	0.02
STEM major	0.41	0.39	0.01	0.59	0.52	0.07***
Experience	10.38	10.54	−0.16	10.87	11.31	−0.44*
HS Certified	0.79	0.75	0.04***	0.81	0.77	0.04***
Education degree	0.54	0.52	0.02	0.56	0.58	−0.02*
Observations	12,760	1630		11,080	1540	

Table 9
Teacher characteristics (Mean) – Makes subject interesting.

Variable	Math			Science		
	No	Yes	Diff	No	Yes	Diff
Female	0.63	0.59	0.03***	0.6	0.55	0.05***
More than Bachelor	0.53	0.5	0.03***	0.58	0.57	0.02
STEM major	0.43	0.4	0.03***	0.57	0.59	−0.02**
Experience	10.87	10.15	0.71***	11.43	10.74	0.69***
HS Certified	0.8	0.78	0.02***	0.82	0.8	0.02***
Education degree	0.52	0.55	−0.03***	0.57	0.56	0.01
Observations	5310	9150		3710	9000	

5.4. Teacher-student sorting

One concern expressed in the literature regards the sorting of students into classroom. For instance, low ability female students may be systematically assigned to female teachers. Therefore, teacher characteristics would no longer be exogenous. This would not be an issue if such sorting mechanisms were based on observable student characteristics, which are controlled for. For instance, if students with low GPA were assigned to female teachers, this would not undermine the above identification strategy as past grades are included as controls (Paredes, 2014). Similarly, if the sorting mechanisms were the same for math and science teacher, it would be taken into account by the student fixed effects. But if the sorting mechanisms were based on student unobservables, different across subjects, and such unobservables were related to the outcome variables and one of the regressors, it would raise an endogeneity problem.

One way to dissipate this concern is to verify how teachers are actually assigned between classes. In the HSLS:09, 9th grade math and science teachers are asked to what extend they agree or disagree with the statement "All or most [math/science] teachers are assigned at least one section of advanced courses", as well as "Advanced courses are assigned to teachers with the strongest [math/science] background". The answers are similar between the two groups. Therefore, this evidence suggests that the teacher sorting mechanism is the same across subject, thus it has been already taken into account by including the student fixed effects.

Similarly, it is possible to argue that students' and parents' sorting behavior is not an issue in this case. First, it seems plausible that students are sorted similarly in math and science classes: if for instance a student or a parent had a preference for female teachers, this would likely be true in both math and science classes, so it would be taken into account in the student fixed effect. Second, among students who agree

Table 10
Teacher characteristics (Mean) – All students can succeed.

Variable	Math			Science		
	No	Yes	Diff	No	Yes	Diff
Female	0.62	0.6	0.02	0.58	0.57	0.01
More than Bachelor	0.51	0.51	0	0.59	0.57	0.02
STEM major	0.41	0.41	0	0.54	0.59	−0.05***
Experience	10.76	10.41	0.36	11.7	10.88	0.82***
HS Certified	0.82	0.78	0.04***	0.83	0.8	0.03**
Education degree	0.5	0.54	−0.04**	0.57	0.56	0.01
Observations	1050	13,380		1030	11,650	

or strongly agree that what they are learning in their science course will be useful for college, 98% of them also agree or strongly agree that their math course will be useful for college. Therefore, since students understand the value of both classes, it is likely that they have put the same amount of effort in searching for the teacher who is the most appropriate for them. Third, information about the placement policy in the schools is provided in the HSLS:09 by the school counselors. Specifically, they are asked about the importance of student/parent choice for 9th grade science/math class. Again, their answers are similar for the two subjects.

Furthermore, the conclusions do not change after controlling for whether the parents think that males are better than females in math or science, and for how much they feel confident in helping their offspring in math/science homework. Therefore, these variables approximate parental efforts (or knowledge) in securing good math and science teachers for their children.

Last, but not least, female and male students are not assigned to different math and science teachers based on observable characteristics. For instance, girls are as likely to be assigned to certified teachers as boys, both in math and science. When there are (small) differences, these appear in both subjects: female students are slightly more likely (1–2 percentage points) to be assigned to a female teacher in math or science. In a symmetric way, female teachers receive students of similar "quality" in math and science: in both subjects, female teachers are 1–2 percentage points more likely to have students who got an A in math or science when they were in 8th grade. To conclude, there is no evidence of different teacher-student sorting mechanisms in math and science.

6. Conclusions

This paper investigates why teacher gender seems to matter for student performances. Using a student fixed effects model, it estimates significant impacts of teacher gender on students' interest and self-confidence in math and science. However, it proves that such effects become indistinguishable from zero once teacher behaviors, expectations and attitudes are controlled for. In particular, how teachers treat boys and girls in the classroom - as well as how they compare males and females in math and science - drive the results.

The empirical analysis also shows that student interest and self-efficacy are substantially affected by teacher ability to make their subject interesting and to create a positive learning environment. These results hold for both male and female students. This is particularly relevant from a policy perspective. Indeed, also given the promising results from interventions aimed at increasing empathy (Okonofua, Paunesku, & Walton, 2016) and reducing gender biases (Carnes et al., 2015), rather than hiring more female teachers or segregating students by gender, training teachers could be more effective in increasing student self-efficacy and interest in STEM. In other words, the aforementioned results highlight the role of teacher quality and effort: what matters primarily in this context are not the role models played by teachers (or the stereotype threats), but the time and skills that instructors put in preparing their lectures and supporting their students.

From a gender perspective, scholars have also been concerned that female students may perform worse than their male counterparts because of low self-confidence. Indeed, in the HSLS:09 sample, female students have lower self-efficacy than their male classmates both in math and science. This study explains how educators could improve this factor though high school teachers.

Finally, I have also verified whether it is possible to identify high-performing teachers from observable characteristics. That is, I have explored whether teachers with more experience, certified to teach in high school, with advanced degrees, and with specific training in Education or STEM are more likely to treat all students equally, to listen and value students' ideas, to have high expectations for all students, to make the subject interesting, and not to have biased gender attitudes. While such characteristics are associated with how teachers compare males and females in math and sciences, the same cannot be said about the other variables. As indicated by the vast literature on the subject (Chetty, Friedman, & Rockoff, 2014a,b; Guarino, Reckase, & Wooldridge, 2015), how to correctly identify, evaluate and incentivize good teachers is still an open question.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.econedurev.2017.09.004.

References

Antecol, H., Ozkan, E., & Serkan, O. (2015). The effect of teacher gender on student achievement in primary school: Evidence from a randomized experiment. *Journal of Labor Economics, 33*(6453), 38.
Apple (2015). Inclusion & diversity. Retrieved from http://www.apple.com/diversity/.
Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review, 95*(2), 152–157.
Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27–41.
Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Valentino, L. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review, 45*, 14–27.
Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., et al. (2015). Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine: Journal of the Association of American Medical Colleges, 90*(2), 221–230.
Carnevale, A. P., Smith, N., & Melton, M. (2011). *Stem.* CEW-Georgetown University1–112.
Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics, 125*(August), 1101–1144.
Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interes, 15*(3), 75–141.
Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593–2632.
Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review, 104*(9), 2633–2679.
Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*(1), 43–54.
College Board. (2016). *SAT: Total group profile report.* .
Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2), 158–165.
Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources, 42*(3), 528–554.
Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences, 114*(23), 5964–5969.
Executive Office of the President. (2012). *Engage to excel: Producing one million additional. College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*.
Filippin, A., & Paccagnella, M. (2012). Family background, self-confidence and economic outcomes. *Economics of Education Review, 31*(5), 824–834.
Fryer, R. G. (2014). Injecting charter shool best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics, 162*(2), 1355–1407.
Fryer, R. G., Jr, & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2*(April), 210–240.
Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student – teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209–224.
Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters. *Economics of Education Review, 29*(6), 911–922.
Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy, 10*(1), 117–156.
Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*(5880), 1164–1165.
Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*(3–4), 153–166.
Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7–8), 798–812.
Hoffmann, F., & Oreopoulos, P. (2009). A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources, 44*(2), 479–494.
Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher. *Labour Economics, 15*(1), 37–53.
Ingels, S. J., Pratt, D. J., Herget, D., Bryan, M., Fritch, L. B., Ottem, R., et al. (2015). *High*

school longitudinal study of 2009 (HSLS : 09) 2013 update and high school transcript data file documentation, Washington, DC. Vol. NCES.
Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., et al. (2011). *High school longitudinal study of 2009 (HSLS:09). Base-year data file documentation. 2011–328* Washington, DC
Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., et al. (2014). *High school longitudinal study of 2009 (HSLS:09) base year to first follow-up data file documentation. 2014–361* Washington, DC
Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615–631.
Koch, A., Nafziger, J., & Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic Behavior and Organization, 115*, 3–17. http://doi.org/10.1016/j.jebo.2014.09.005.
Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics, 92*(10–11), 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*.
Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes : Implications for theory and practice. *Studies in Higher Education, 27*(1), 37–41.
Mechtenberg, L. (2009). Cheap talk in the classroom: how biased grading at school explains gender differences in achievements, career choices and wages. *The Review of Economic Studies, 76*(4), 1431–1459.
Muralidharan, K., & Sheth, K. (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources, 51*(2), 269–297.
National Science Foundation. (2015). *Women, minorities, and persons with disabilities in science and engineering. Special Report NSF 15-311. Arlington, VA*.
NCES (2016). Mathematics literacy: Gender. Retrieved February 13, 2017, from https://nces.ed.gov/surveys/pisa/pisa2015/pisa2015highlights_5c.asp.
OECD. (2015). *The ABC of gender equality in education.* PISA: OECD Publishing.
Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences, 113*(19), 5221–5226. http://doi.org/10.1073/pnas.1523698113.
Papageorge, N. W., Gershenson, S., & Kang, K. (2016). Teacher expectations matter. *IZA Discussion Paper, 10165*.
Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review, 39*, 38–49.
Patterson, B., & Kobrin, J. (2012). *The SAT and SAT subject tests: Discrepant scores and incremental validity.* College Board.
Pollack, E. (2013). *Why are there still so few women in science?* October 3New York Times.
Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review, 29*(6), 901–910.
Schneeweis, N., & Zweimüller, M. (2012). Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review, 31*(4), 482–500.
Weinberger, C. J., & Leggon, C. B. (2004). *Just Ask! Why surveyed women did not pursue IT courses or careers.* IEEE Technology And Society Magazine, Spring28–35.
Winters, M. A., Haight, R. C., Swaim, T. T., & Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review, 34*, 69–75.
Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources, 48*(3), 545–595.