Economics of Education Review 31 (2012) 54–67

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

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

The effect of teacher–student gender matching: Evidence from OECD countries

Insook Cho*

School of Technology Management, Ulsan National Institute of Science and Technology, #100 Banyeon-ri, Eonyang-eup, Ulju-gun, Ulsan 689-798, South Korea

ARTICLE INFO

Article history:
Received 21 November 2010
Received in revised form 26 January 2012
Accepted 3 February 2012

JEL classification:
I20
J24

Keywords:
Teacher–student gender matching
Academic achievement
Fixed effect model
TIMSS

ABSTRACT

While some educators argue that teacher–student gender matching improves student performance, there is little empirical evidence to support this hypothesis. This paper assesses the impact of teacher–student gender matching on academic achievement across fifteen OECD countries using data from the Trends in International Mathematics and Science Study (TIMSS). One attractive feature of TIMSS is that it provides information on test scores and teacher characteristics, including gender, for both math and science thereby allowing for student fixed effects estimation. The results provide little support for the conjecture that students benefit from teacher–student gender matching.

1. Introduction

Gender difference in the educational outcomes of young children is a concern for both parents and educators. Recent explanations addressing the gender gaps observed in children's academic achievement often emphasize the important role of teachers. For instance, numerous studies have claimed that teacher–student interactions benefit one sex at the expense of the other.[1] While there is little evidence of systematic discrimination against students by teachers of the opposite sex, empirical studies indicate that students do experience different interactions with male and female teachers.[2] There are many theories concerning how and why educational outcomes are influenced by the gender of students and teachers. It has been suggested that a teacher of the same gender may positively influences students' outcomes by communicating with them more effectively, by showing high expectations for enhanced performance (Pygmalion effect; Braun, 1976; Rosenthal & Jacobson, 1968), or by being a good role-model (role-model effect; Almquist & Angrist, 1971; Basow & Howe, 1980). Furthermore, a teacher of the opposite-gender may have a detrimental effect on students' outcomes by strengthening the pressure of negative stereotypes (Steele, 1997).

Empirical studies similarly provide mixed results concerning the effect of teacher's gender on educational outcomes. For instance, Dee (2007) finds that teacher–student gender interactions positively influence the teachers' subjective evaluation of student behavior and performance. While some studies find that female faculty

* Tel.: +82 52 217 3114; fax: +82 52 217 3109.
E-mail address: icho@unist.ac.kr

[1] Some critics posit that boys lag academically because female teachers are unable to meet their learning needs (Kleinfeld, 1998; Riordan, 1999; Sommers, 2000). Other critics claim that girls are not reaching their potential because teachers favor boys by giving more positive feedback to boys or having more frequent social interactions with boys (AAUW, 1992; Becker, 1981).

[2] See Brophy (1985) for an extensive literature review.

0272-7757/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2012.02.002

have a positive effect on the educational attainment and course selection of female students in high schools and colleges (Nixon & Robinson, 1999; Robst, Keil, & Russo, 1998), other studies have concluded that there is little evidence corroborating this hypothesis (Ehrenberg, Goldhaber, & Brewer, 1995; Krieg, 2005). Dee (2007) maintains that these contradictory results may partly reflect the fact that reduced-form estimates are biased by unobserved student traits that are correlated with the gender of the teacher. For example, school administrators may disproportionately assign low-achieving students to female teachers because they believe that female teachers more effectively address the needs of these students. This type of nonrandom assignment would result in a negative correlation between teacher gender and unobservable student ability, therefore biasing the teacher gender coefficient. To address this type of potential bias, Dee (2007) uses a student fixed effects model, which differences out unobservable student traits. Using data for eighth graders in the United States, Dee (2007) finds that teacher's gender is positively correlated with students' test scores, teacher's perceptions of student performance, and student engagement with the teacher's subject.[3]

Despite the inconsistent evidence regarding the impact of teacher's gender on students' academic outcomes, substantial efforts have been made to recruit and retain female teachers in specific subjects (math, science, and engineering) and to increase the proportion of male teachers, especially at the elementary school level. While these teacher hiring policies are based on the belief that students benefit from same-gender teachers, they also cause reason for concern. First, aggressive attempts to recruit teachers of a specific gender may lead to lower quality hires. Second, it is possible that teachers of one gender might improve academic achievements of students of the same gender while lowering the academic achievement of opposite gender students. Focusing on the second possibility, this paper presents empirical evidence gathered across multiple countries pertaining to the impact of teacher–student gender matching on both male and female student test scores. Using a student fixed effect model similar to that of Dee (2007), this paper examines whether teacher's gender impacts student academic achievement in multiple countries. The evidence presented in this paper provides no strong support for the hypothesis that teacher's gender has a significantly positive impact on students' test scores.

The contributions of this paper to the literature are twofold. First, the empirical evidence of this study suggests that the advantages of teacher–student gender matching on student academic achievement has diminished over the last two decades in the United States, especially in math and science subjects. Contrary to Dee's (2007) findings, this study demonstrates that assignment to a same gender teacher neither improves nor diminishes students' test scores in the United States. The contrasting outcomes of the two studies may stem from the use of different data sources and different identifying assumptions. The data used in this study are from the four waves of the Trends in International Mathematics and Science Study (TIMSS; 1995, 1999, 2003, 2007) for lower secondary school students whereas Dee uses 1988 data for eighth graders from the National Education Longitudinal Study of 1988 (NELS:88). More significantly, the two studies use somewhat different identification strategies. Observations of a pair of subjects are used to difference out unobservables, but the subject pairs differ across studies. The structure of NELS dictates that Dee must pair a math or science subject with an English or history subject. This form of subject pairing is necessary because teacher data is only available for either math or science and either English or history. In contrast, this study employs only one subject pair – math and science – for all students. Therefore, Dee's identifying assumption is that unobservable student specific characteristics (e.g., ability and preference) are the same across the observed math/science and English/history subject pairs. Conversely, the identifying assumption for this study is that unobservable student specific characteristics are the same for math and science. To the extent that math and science are better controls for one another than math or science paired with English or history (see Gardner & Hatch, 1989), the data of this study serves to more adequately remove student specific unobservables.[4]

Second, this study estimates the teacher–student gender matching effects in multiple countries in order to test the universal teacher gender effect hypothesis. If, as educational psychologists expect, the educational gains or losses of students matched with a same gender teacher result from the teacher–student interaction, then observations of a significant teacher gender effect across multiple countries would be expected. However, as reported in this paper, the results from 15 sample countries cast substantial doubt on this view. In eight of 15 countries, teacher's gender has no significant impact on student test scores. Of the remaining seven countries, assignment to a same gender teacher positively correlates with boys' test scores in four countries (Canada, Japan, Portugal, and Spain) and positively correlates with girls' test scores in three countries (France, Greece, and Sweden). Moreover, this paper provides evidence that these observed positive correlations between teacher's gender and student achievement can be explained by differences in teacher quality. Essentially, the empirical results of this paper contradict the hypothesis that assignment to a same gender teacher positively affects student performance. Coordinately, several studies which adopt the fixed effect frameworks provide consistent empirical evidence on the issue.[5] To reiterate,

[3] This study finds that teacher gender has no significant impact on student academic achievement in the United States. In contrast, Dee (2007) reports evidence supporting the teacher–student gender matching effect among eighth graders in the United States. He finds that assignment to a same gender teacher improves boys' test scores in science by 0.049 standard deviations and girls' test scores in history by 0.077 standard deviations. In addition, he finds that assignment to a male math teacher improves both boys' and girls' test scores by 0.076 and 0.065 standard deviations, respectively.

[4] The comparative review of the results of this study and Dee (2007)'s is more extensively discussed in the last three paragraphs of Section 4.2.

[5] Using the Progress in International Reading Literacy Study (PIRLS 2001) and the three waves of the Trends in International Mathematics

the empirical evidence presented in this paper provides no support for the hypothesis that students and teachers sharing the same gender impacts student performance on standardized tests.

The remainder of the paper is organized as follows. Section 2 describes the data used in this paper, while Section 3 details the first difference (FD) framework used to estimate the effect of teacher's gender on student outcomes. The empirical results are presented in Section 4 through comparison with existing literature. Section 5 presents robustness checks for the first difference estimates of this paper. Conclusions are given in Section 6.

2. Data: TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is an international study of student academic achievement in math and science. The study, conducted by the International Association for the Evaluation of Educational Achievement (IEA), assessed students' math and science skills in 1995, 1999, 2003, and 2007. It also surveyed students, teachers and school administrators to gain in-depth information about the various factors that contribute to educational outcomes. Using the TIMSS is beneficial because the analysis can be extended beyond a single country. Another important aspect of the TIMSS is that it reports student–teacher data for two subjects for each student. The matched pair structure of the data means that student fixed effects models can be adopted to explore the impact of teacher–student gender matching on student outcomes, at least for the junior high school sample.

Three waves of the TIMSS (1995, 2003, 2007) comprise student achievement data for two different grade groups, while the second wave of the TIMSS (1999) comprises data for a single grade group. The 1995 TIMSS, the largest and most complex international study among the four waves, assessed and surveyed students in two grade groups. The first set of scores is for students enrolled in the two adjacent grades that contain the largest proportion of nine year olds – third and fourth graders in most countries, and the second set of scores is for students enrolled in the two adjacent grades that contain the largest proportion of thirteen year olds – seventh and eighth graders in most countries.[6] While the 2003 and 2007 TIMSS both use the 1995 definition to target the two adjacent grades containing the most nine and thirteen year olds, only students in the upper of the two grades were tested – fourth and eighth graders in most countries. In contrast, the 1999 TIMSS includes only one age group in a single grade. Again, the 1999 TIMSS uses the 1995 definition to target the two adjacent grades containing the most thirteen year olds, but only students in the upper of the two grades were tested – eighth graders in most countries. For expositional ease, this paper refers to the sample of seventh and eighth graders in 1995 and the sample of eighth graders in 1999, 2003 and 2007 as lower secondary school students.

While almost all the lower secondary school students are taught math and science by different teachers, all third and fourth graders in the TIMSS receive math and science instruction from a single teacher. The single teacher structure of elementary level math and science instruction does not allow for the first difference estimation of the impact of teacher–student gender matching on student test scores because it is not possible to include student fixed effects. This paper, therefore, focuses on the gender matching effect at the lower secondary school level.

The 1995, 1999, 2003 and 2007 TIMSS all survey lower secondary school students from 41, 38, 48 and 50 countries, respectively. The sample for this study, however, is restricted to 15 OECD countries. This restriction is implemented for the following three reasons. First, the sample is restricted to OECD countries in which lower secondary school participation is almost universal. This is important because if a large proportion of students drop out before grade eight (e.g., Turkey), then the sampled students in the TIMSS would not represent the national population of those countries. Second, 11 OECD countries are excluded from the sample because a large proportion of teachers did not complete their surveys which results in the loss of a large number of observations in these countries.[7] Third, Italy was excluded from the sample because lower secondary school students in Italy are taught math and science by a single teacher.

Imposing the above restrictions yields a sample of 202,644 lower secondary school students from 15 OECD countries. Excluding students who did not report their sex and students whose teachers did not submit their information further reduces the sample by 1167, thus resulting in a final sample of 201,477 students and 402,954 observations. Due to the complex sample design of the TIMSS 1995, the sample of this study includes 46,543 and 1660 secondary school students in the seventh and ninth grades, respectively, as well as 153,274 students in the eighth grade. Summary statistics are reported in Table 1.

The TIMSS test scores used in all analyses are standardized within test book across all TIMSS participants to have a mean of 0 and a standard deviation of 1. As exhibited in Table 1, the average test scores for many sample countries are higher than the international mean 0, because students from OECD countries tend to perform better on internationally standardized tests than students from non-OECD countries. Taking into account that teacher qualifications

and Science Study (TIMSS; 1995, 1999, 2003), Ammermuller and Dolton (2006) find no evidence of a positive gender matching effect among eighth graders in the United States. Ouazad (2008), employing the Early Childhood Longitudinal Study (ECLS 1998), also finds that teacher's gender does not significantly impact academic achievement for elementary school students in the United States. Holmlund and Sund (2008) find no evidence that teacher–student gender matching improves student outcomes among upper secondary school students in Sweden. In accordance, a study in Germany by Neugebauer, Helbig, and Landmann (2011) found no evidence that boys or girls benefit from having a same-sex teacher.

[6] In Sweden, the 1995 TIMSS surveyed lower secondary school students from three adjacent grades, grades seven, eight and nine. The sample of this study, therefore, includes 1660 students in the ninth grade from Sweden.

[7] Appendix Table 1 reports the FE estimates for teacher–student gender matching for these 11 OECD countries. The analysis of other OECD countries provides little support for the universal teacher–student gender matching effect hypothesis.

Table 1
Summary statistics.

Country	Sample size	Test scores		% of female students	% female teachers: math	% female teachers: science
		Math	Science			
	(1)	(2)	(3)	(4)	(5)	(5)
Canada	22,516	0.11	0.11	49.9%	45.1%	42.1%
Czech Republic	14,627	0.44	0.59	50.2%	78.9%	72.1%
Finland	2908	0.30	0.45	50.3%	59.5%	62.3%
France	5170	0.14	−0.25	49.6%	48.6%	53.4%
Greece	7064	−0.37	−0.26	48.9%	33.3%	42.9%
Hungary	15,764	0.42	0.54	50.5%	82.1%	74.8%
Japan	23,864	0.84	0.57	49.1%	30.0%	18.8%
Netherlands	9133	0.52	0.53	49.9%	28.4%	22.5%
Norway	13,449	−0.09	0.07	49.8%	41.6%	41.9%
New Zealand	13,605	0.00	0.15	48.8%	48.9%	50.1%
Portugal	6391	−0.64	−0.45	50.3%	66.7%	77.7%
Slovak Republic	14,303	0.38	0.37	50.5%	80.8%	74.6%
Spain	7123	−0.33	−0.05	51.3%	42.1%	46.1%
Sweden	16,680	0.18	0.31	49.1%	47.7%	45.3%
U.S.A.	28,926	0.14	0.30	51.1%	64.4%	54.5%

Notes: Test scores are internationally standardized to a mean of 0 and standard deviation of 1. In most sampled countries, the average test scores are higher than the international mean 0, because students in developed countries tend to perform better than their counterparts in developing counties.

and classroom resources may have a discernible impact on student outcomes (Clotfelter, Ladd, & Vigdor, 2007), all of the models used in this paper include a set of observable teacher, classroom, and school characteristic variables: the teacher's education level measured by 5 categorical dummies, the teacher's teaching experience measured by 10 categorical dummies, class sizes, the proportion of students having been the victim of theft at school, the proportion of students having been harmed by others at school, and the dummy variables for regions in which schools are located.[8] The TIMSS also provides indicators for whether the teacher is state-certified in the subject and whether the teacher is trained to teach primarily in math, in science, or in other subjects. The indicators for teacher certification and whether the teacher instructs his/her main field subjects are also included in the models where possible. All models include country-by-year and country-by-grade dummies.[9] In addition to the teacher and classroom control variables, the OLS models also include a basic set of socioeconomic controls including parental education, student's and parent's immigrant status, whether a student lives with both mother and father, the number of books in the home, and household size.

It is important to mention the basic sample design of the TIMSS, which is generally referred to as a two-stage stratified cluster sample design. In the first stage, the TIMSS selects 150 or more nationally representative schools depending on location, number of pupils, whether the school is public or private, and other school characteristics. All lower secondary school students in the sampled schools take internationally standardized math and science tests. In the second stage, a single classroom in the sampled school is randomly selected, and students in that classroom are included in the TIMSS data set. Since the sampled students are all from the same class and have the same pair of math and science teachers, adding dummy variables that identify individual teachers or classrooms causes a perfect multicollinearity with the teacher gender variable. In contrast to Dee (2007), the models administered in this study do not include teacher or classroom fixed effects.

3. Specification: first difference (FD) model

To identify the effect of teacher gender on student outcomes, a first difference (FD) model is used. This estimation strategy was first proposed by Dee (2007) as a simple variation of the research design implemented to estimate the return to education using samples of twins (Ashenfelter & Krueger, 1994). I assume that the test score of student i in country c for subject j (Y_{ijc}; $j = M$ for math, $j = S$ for science) is a function of whether the subject j teacher shares the same-gender with student i in country c ($SAME_{ijc}$), observable teacher and classroom characteristics (Z_{ijc}), observable characteristics of student i (X_{ic}), unobservable student fixed effect (μ_{ic}), and a mean-zero error term (ε_{ijc}). Eqs. (1) and (2) are simple linear models that explain student's test

[8] The set of control variables used in Dee's (2007) study include a dummy variable that identifies whether the student shares the teacher's race/ethnicity, whether teacher is state-certified in the subject, teacher experience measured by 10 categorical dummies, number of students in the class, and the percentage of students in the class who are limited English proficiency. The TIMSS does not provide information on teacher's race/ethnicity and percentage of students with limited language proficiency.

[9] Irvine (1986) argues that the grade level of students can influence the outcomes of teacher–student interaction because the student's interaction/communication process changes as students move up to higher grade levels. For example, students in lower grades tend to receive more frequent and individualized feedback from their teachers because students are more likely to interact with their teachers in small group settings. Conversely, students in upper grades are more likely to be involved in more independent study, and consequently, have fewer interactions with the teacher. Although the 1995 wave of TIMSS surveyed secondary school students from two (three for Sweden) different grades, it is doubtful whether any significant difference can be observed in gender matching effect between seventh graders and eighth graders. The data show that there is no significant difference in the gender matching effect between students in two grades in most countries.

scores in math and science, respectively, with student, teacher, and classroom characteristic variables.

$$Y_{iMc} = \alpha_{Mc} + \beta_c SAME_{iMc} + \lambda_c Z_{iMc} + \delta_c X_{ic} + \mu_{ic} + \varepsilon_{iMc} \quad (1)$$

$$Y_{iSc} = \alpha_{Sc} + \beta_c SAME_{iSc} + \lambda_c Z_{iSc} + \delta_c X_{ic} + \mu_{ic} + \varepsilon_{iSc} \quad (2)$$

A positive and significant coefficient for the teacher gender variable ($SAME_{ijc}$), β_C, would suggest that students matched with a same gender teacher score higher than those matched with an opposite gender teacher. However, it is of concern that the coefficient for teacher gender using this simple model may be biased because of the possible correlation between teacher gender ($SAME_{ijc}$) and student fixed effect (μ_{ic}). For example, if students in lower academic tracks are more likely to be assigned to female teachers, this nonrandom assignment creates a negative correlation between teacher gender and unobservable student ability and causes a bias in the coefficient for the gender matching variable. Differencing out Eqs. (1) and (2) yields the first difference model, Eq. (3), whose coefficients are not compromised by possible correlation between teacher gender variable ($SAME_{ijc}$) and unobserved student fixed effects (μ_{ic}).

$$(Y_{iMc} - Y_{iSc}) = (\alpha_{Mc} - \alpha_{Sc}) + \beta_c(SAME_{iMc} - SAME_{iSc}) + \lambda_c(Z_{iMc} - Z_{iSc}) + (\varepsilon_{iMc} - \varepsilon_{iSc}) \quad (3)$$

The identifying assumption is that unobservable student specific characteristics (μ_{ic}) are the same for math and science. That is, the unobservable student's fixed effects (e.g., ability, preferences) are assumed to have the same impact on both math and science test scores. Finding a positive coefficient for the gender matching variable ($SAME_{iMc} - SAME_{iSc}$) would indicate that assignment to a same gender teacher improves students' test scores. This baseline model estimates the gender matching effect separately by gender as in Dee's (2007) study. The benefit of Dee's (2007) strategy is that one can study how teacher–student gender interaction affects student's test scores for boys and girls separately. Moreover, the (student) gender specific model allows all the observable teacher/classroom characteristics, including teacher gender, to have a differing impact on students' outcomes by student gender. The estimation results from this baseline model are comparable to Dee's (2007).

The baseline model above assumes that the effect of teacher gender in math is equal to that in science. However, it would be reasonable to allow the effect of a teacher's gender to differ across subjects. For example, it is possible that the role-model effect is bigger in math than in science. The following specification allows for a subject specific teacher–student gender matching effect.

$$Y_{iMc} = \alpha_{Mc} + \beta_{Mc} SAME_{iMc} + \lambda_c Z_{iMc} + \delta_c X_{ic} + \mu_{ic} + \varepsilon_{iMc} \quad (4)$$

$$Y_{iSc} = \alpha_{Sc} + \beta_{Sc} SAME_{iSc} + \lambda_c Z_{iSc} + \delta_c X_{ic} + \mu_{ic} + \varepsilon_{iSc} \quad (5)$$

$$(Y_{iMc} - Y_{iSc}) = (\alpha_{Mc} - \alpha_{Sc}) + \beta_{Mc} SAME_{iMc} + \beta_{Sc}(-SAME_{iSc}) + \lambda_c(Z_{iMc} - Z_{iSc}) + (\varepsilon_{iMc} - \varepsilon_{iSc}) \quad (6)$$

If the math estimate (β_{Mc}) is statistically different from the science estimate (β_{Sc}), that would indicate that the effect of gender matching in a math classroom is different from that in a science classroom. The subject specific gender matching effects are estimated separately for boys and girls.

In order to control for time trends and grade level differences, all of the models include country-by-year and country-by-grade dummies. Since there are two observations per student, the estimation results from the first difference (FD) model and from the fixed effects (FE) model are equivalent. Therefore, all models are estimated using fixed effects. In all models, the standard errors are clustered at the school level.

4. Results

4.1. Baseline results

Table 2 reports the point estimates for the effect of assignment to a same gender teacher on student performance in 15 OECD countries. Columns 1 and 3 report the

Table 2
The gender matching effect: the OLS and the FE estimates.

	Boys		Girls	
	OLS (1)	FE (2)	OLS (3)	FE (4)
Canada	0.06***	0.02	−0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Czech Republic	0.09***	0.00	−0.05**	0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Finland	−0.01	0.01	0.01	0.01
	(0.03)	(0.02)	(0.03)	(0.02)
France	0.00	−0.02	0.06	0.02
	(0.04)	(0.02)	(0.04)	(0.02)
Greece	−0.04	−0.01	0.03	0.04**
	(0.03)	(0.01)	(0.02)	(0.01)
Hungary	−0.01	0.02*	0.00	0.00
	(0.02)	(0.01)	(0.02)	(0.01)
Japan	−0.02	0.02	0.05**	0.02
	(0.02)	(0.01)	(0.02)	(0.01)
Netherlands	−0.06	−0.01	0.07**	0.01
	(0.04)	(0.01)	(0.04)	(0.01)
Norway	−0.03	−0.02	0.02	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
New Zealand	−0.02	0.00	0.08***	0.02
	(0.03)	(0.02)	(0.03)	(0.02)
Portugal	−0.02	0.02	0.02	0.02
	(0.03)	(0.02)	(0.03)	(0.02)
Slovak Republic	0.05	−0.01	0.00	0.01
	(0.03)	(0.01)	(0.03)	(0.02)
Spain	−0.04	0.12**	0.04	−0.01
	(0.04)	(0.05)	(0.04)	(0.06)
Sweden	0.01	−0.01	0.01	0.02**
	(0.02)	(0.01)	(0.02)	(0.01)
U.S.A.	0.02	0.02	−0.02	0.01
	(0.02)	(0.01)	(0.02)	(0.01)

Notes: Robust standard errors, adjusted for clustering on school level, are in parentheses. All models include teacher/classroom controls listed in the data section. Country-by-year dummies as well as country-by-grade dummies are included in all models. The OLS estimates are from a pooled regression model with subject fixed effect. Test scores are standardized to an international mean of 0 and a standard deviation of 1.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

OLS estimates[10] for boys and girls, respectively, in the sample countries. The OLS models control for teacher and classroom characteristics (teacher's education, teacher's experience, and class sizes) as well as for observable student characteristics (parental education, immigrant status, the number of books at home, household size, and whether a student lives with both mother and father). The OLS results indicate that students' test scores are positively correlated with the gender of the teacher only in five out of 15 countries. In Canada, boys matched with male teacher achieve higher test scores than boys matched with a female teacher. In Japan, Netherlands and New Zealand, girls matched with a female teacher achieve higher test scores. In Czech Republic, not only boys but also girls score significantly higher when they are assigned to a male teacher.

As previously discussed, one of the main concerns regarding the OLS estimates is the failure to address the endogeneity of observed gender matches. For instance, it is possible that school administrators tend to assign under-achieving students to female teachers because they believe that female teachers are more effective in helping these students. In this case, non-random student assignment induces negative correlation between assignment to female teachers and achievement resulting in downward biased teacher–student gender matching estimates. Comparing OLS and FE estimates provides a useful indication of the possible correlation between teacher gender and unobservable student characteristics. Referring to Table 2, the OLS estimates for the Czech Republic indicate that both boys and girls assigned to female teachers achieve lower test scores by 0.09 and 0.06 standard deviations, respectively, compared to students assigned to male teachers. In contrast, the FE estimates reported in columns 2 and 4 show a small and imprecise impact of assignment to a same gender teacher on both boys' and girls' test scores when controlling for student fixed effect. This suggests that observed female teachers' negative effect in the OLS estimates may reflect non-random sorting. In other words, the OLS estimates may suffer from potential bias due to the possible correlation between teacher gender and unobservable student characteristics. The FE estimation strategy addresses this concern by controlling for student fixed effects.

Therefore, the key results of this study on the teacher–student gender matching effect are the fixed effect (FE) estimates using Eq. (3). The FE estimates for boys and girls are reported in columns 2 and 4, respectively. The FE results do not provide strong support for the hypothesis that teacher–student gender matching significantly impacts students' educational outcomes. The estimates indicate that, in 11 of 15 countries, assignment to a class taught by a teacher of the same gender has no significant impact on students' test scores. Only in four countries (Greece, Hungary, Spain and Sweden) do students matched with a same gender teacher have significantly higher test scores than students matched with an opposite gender teacher. In Greece and Sweden, girls matched with a female teacher achieve higher test scores by 0.03 and 0.02 standard deviations, respectively, compared to girls matched with a male teacher at the 5 percent level. In Spain, assignment to a male teacher improves boys' test scores by 0.12 standard deviations at the 5 percent significant level. In Hungary, assignment to a male teacher raises boys' test scores by 0.02 standard deviations at the 10 percent level. Collectively, the results from baseline models provide limited evidence of a teacher–student gender matching effect across numerous OECD countries.

4.2. Subject specific effects

Table 2 reports the FE estimates based on models which assume that the teacher–student gender matching effect is the same for math and science. However, it is also reasonable to allow the gender matching effect for math to differ from that of science. For example, the role-model effect may be more or less distinctive in math than in science. Eq. (6) is a generalization of Eq. (3) and, therefore, allows for subject specific teacher–student gender matching effects. Table 3 presents the subject specific effect of gender matching in 15 countries using Eq. (6). Only two of the 15 estimates for boys and three of the estimates for girls are statistically significant at the 5 percent level. In two other countries, the estimates for boys are statistically significant at the 10 percent level. Taken as a whole, the results in Table 3 provide little support for the notion that teacher–student gender matching improves test scores for either girls or boys.

Despite the generally insignificant gender matching estimates, it is worth briefly discussing the statistically significant findings for seven countries. The results in columns 1 and 2 indicate that teacher gender is positively correlated with boys' test scores at the 5 percent level in two countries (Canada, Spain), and at the 10 percent level in two countries (Japan, Portugal). In addition, the subject specific FE estimates suggest that the significant impacts on boys' test scores are driven predominantly by male science teachers (in three of four countries). In Canada, boys matched with a male science teacher achieve higher test scores by 0.06 standard deviations at the 5 percent level. In Japan and Portugal, assignment to a same gender teacher raises boys' test scores by 0.04 standard deviations in science subjects at the 10 percent level. In Spain, boys achieve higher test scores both in math and science by 0.12 and 0.13 standard deviations, respectively, when matched with a male teacher. Columns 3 and 4 depict the gender matching effect in math and science, respectively, for girls. The results indicate that assignment to a same gender teacher has a positive and significant impact in only three countries (France, Greece, Sweden). Contrary to the results for boys, the significant and positive impacts on girls' test scores are mostly driven by female math teachers (in two of three countries). In France and Greece, assignment to a female teacher raises girls' math test scores by 0.09 and 0.06 standard deviations, respectively, at the 5 percent level. In Sweden, teacher–student gender matching improves girls' science test scores by 0.03 standard

[10] The OLS estimates reported in columns 1 and 3 are from a pooled regression model with subject fixed effect. These are comparable to the OLS estimates in Dee's (2007) study.

Table 3
The subject specific gender matching effect: the FE estimates.

	Boys		Girls	
	Math (1)	Science (2)	Math (3)	Science (4)
Canada	−0.03	0.06**	0.04	−0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Czech Republic	0.00	0.00	0.01	0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Finland	0.05	−0.01	0.01	0.02
	(0.04)	(0.02)	(0.04)	(0.02)
France	−0.03	−0.01	0.09**	−0.02
	(0.03)	(0.02)	(0.04)	(0.02)
Greece	0.01	−0.02	0.06**	0.00
	(0.03)	(0.02)	(0.03)	(0.02)
Hungary	0.03	0.01	−0.01	0.00
	(0.03)	(0.01)	(0.02)	(0.01)
Japan	0.00	0.04*	0.02	0.02
	(0.01)	(0.02)	(0.01)	(0.02)
Netherlands	−0.04	0.00	0.02	0.01
	(0.03)	(0.01)	(0.03)	(0.01)
Norway	−0.02	−0.01	−0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
New Zealand	−0.01	0.00	0.01	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Portugal	0.00	0.04*	0.03	0.01
	(0.03)	(0.02)	(0.03)	(0.02)
Slovak Republic	−0.05	0.01	0.01	0.01
	(0.03)	(0.01)	(0.03)	(0.01)
Spain	0.12**	0.13**	−0.01	0.00
	(0.05)	(0.06)	(0.06)	(0.06)
Sweden	−0.01	−0.01	0.02	0.03**
	(0.02)	(0.01)	(0.02)	(0.01)
U.S.A.	0.01	0.03	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.01)

Notes: Robust standard errors, adjusted for clustering on school level, are in parentheses. All models include teacher/classroom controls listed in the data section. Country-by-year dummies as well as country-by-grade dummies are included in all models. Test scores are standardized to an international mean of 0 and a standard deviation of 1.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

deviations at the 5 percent level. The results depicted in Table 3 affirm that, depending on the subject, teacher–student gender matching may differently impact students' academic achievements. Nonetheless, the results from the flexible models do not provide convincing evidence to substantiate a universal teacher–student gender matching effect across 15 OECD countries.

Subsequently, this paragraph introduces several possible explanations as to why the findings of this study differ from those of Dee (2007), who finds significant and positive effects of a same gender teacher on student test scores in the United States. The divergence between the two studies is a result of differences in data sources, time periods, and identification strategies. First, this study utilizes the four waves of an international study (TIMSS 1995, 1999, 2003, 2007) which surveyed lower secondary school students from more than sixty countries including the United States. Dee's (2007) study, on the other hand, utilizes a national study which surveyed eighth graders from only the United States (NELS:88). While both data sets collect information from nationally representative samples of students in the United States, the designers of the TIMSS should make additional effort to construct data that is comparable across multiple countries. Resultantly, the survey items of the TIMSS would then be carefully designed to represent the curriculum of all participating countries and to avoid any bias towards or against particular countries.[11] The variations in survey item design may cause two data sets to evaluate student characteristics (e.g., standardized test scores) or to measure teacher characteristics somewhat differently. This in turn may have led to the divergence in the estimation results.

The second explanation pertains to the difference in the time period for each survey. The data sets used in Dee's (2007) paper were collected in 1988, while the data sets used in this study were collected every four years from 1995 until 2007. As Appendix Tables 2 and 3 illustrate, over the past two decades, there have been substantial changes

[11] Consideration of international comparability results in some noticeable differences in survey items between the two data sets. One example is the standardized tests used in the two data sets. The TIMSS tests include questions in free-respond format which requires students to generate and write their own answers. On the contrary, the NELS tests are all multiple choice tests that fit into 1.5 h of testing time. The distribution of test items also differs between the two data sets. While the TIMSS math test items are distributed relatively evenly across content areas (arithmetic 34%, algebra 18%, geometry 15%, data analysis and probability 21%, others 12%), the NELS math test items are more concentrated in certain content areas (arithmetic 48%, algebra 28%, geometry 10%, data analysis and probability 10%, others 4%).

in teacher characteristics in the United States. The most noticeable change is an increase in the proportion of female teachers in math and science subjects between 1988 and 1995. Another important change is the significant reduction of the gender gap in education levels between male and female teachers. In 1988, male teachers were more likely than female teachers to hold Master's or other professional degrees. However, by 1995 the education level of female teachers became essentially identical to that of male teachers. Finally, the proportion of teachers who are state-certified increased considerably between 1988 and 2003.[12] While one in five math/science teachers were not state-certified in 1988, the data indicates that by 2003 almost all math and science teachers were state-certified. Overall, a comparative review of the NELS and the TIMSS data denotes an extensive increase in the proportion of female teachers in math and science classes along with consistent improvement in teacher quality indicated by education level and certification status. The change in teacher characteristics over time may have affected the teacher–student gender interaction patterns and resulted in the divergence in estimation results between the two studies.

The third explanation for the deviations in the two studies is the difference in identification strategies. This study works under the identification assumption that unobserved student fixed effect (e.g., ability, preference) in math is identical to that in science. On the contrary, Dee's (2007) identification assumption is that student unobservable fixed effect in math or science is identical to that in reading or history. The circumstances of the NELS:88 required Dee (2007) to employ that particular identification assumption as it provides teacher information for only one of four subject matches as follows: math/reading, math/history, science/reading, and science/history. Considering that the skill set required to do well in math is more closely related to that for science than that for reading or other social science subjects, the identification strategy of this study removes the student fixed effect more adequately than that for Dee's.

5. Robustness checks

While the fixed effect (FE) estimates in Tables 2 and 3 provide little evidence supporting a significant academic benefit for teacher–student gender matching in most sample countries, they do indicate that students matched with a same gender teacher achieve higher test scores than students matched with an opposite gender teacher in a small number of countries. The student fixed effect model of this study aims to avoid the potential bias in the coefficient for teacher gender due to the correlation between teacher gender and unobservable student traits. Nevertheless, there are several reasons why the FE estimates in this study may still be biased. The alternative explanations include differences in teacher quality, unobservable teacher/classroom characteristics, or school polices. While it is unfeasible to test all alternative explanations for the potential bias in the FE estimates of this study with the given data set, this section provides evidence that in some countries, differences in teacher quality may explain the observed positive correlation between teacher gender and student achievement.

5.1. Subject-matter knowledge and teacher effectiveness

There is little doubt that the effectiveness of teachers is the single most important factor affecting students' academic growth (Sanders & Horn, 1998). The importance of teacher's subject-matter knowledge is an area of growing interest in current dialog concerning teacher effectiveness. Naturally, teachers prepared to teach mathematics should provide more effective teaching in mathematics classes compared to teachers who trained to teach other subjects. The literature establishing a positive relationship between a teacher's subject-matter knowledge and increased student achievement is voluminous. For example, Goldhaber and Brewer (2000) find that students matched with teachers with a B.A. or M.A. degree in mathematics had greater academic achievement gains in mathematics. Studies using school-level data find a positive correlation between student achievement and teacher's certification status or teacher's scores on certification tests which measure the teacher's subject matter knowledge (Andersson, Johansson, & Waldenstrom, 2011). Further studies reveal that teacher's content knowledge levels matter as well. For example, undergraduate coursework in a particular subject (Monk, 1994), number of credits taken in that subject at college (Begle & Geeslin, 1972), or teacher's undergraduate performance (Kukla-Acevedo, 2009) all positively correlate with student achievement. Existing studies provide strong evidence that students tend to have increased gains in academic achievement when they are assigned to teachers who are teaching their main-field subjects (Monk, 1994; Rowan, Chiang, & Miller, 1997). These findings illustrate that the amount of education training as well as the subject matter content strongly influence teachers' demonstrated skills and effectiveness.

Ideally, teachers should be assigned to the subject they prepared to teach, as opposed to a subject out of their main field. However, as illustrated in Table 4, a large proportion of teachers in many countries are assigned to teach subjects out of their field of study. Table 4 presents various indicators for teacher's educational background and level of math and science subject matter knowledge. The indicators include: proportion of teachers with B.A. or graduate degrees, proportion of teachers with certification, proportion of teachers who majored in math and science, and proportion of students instructed in math and science by a single teacher. According to Table 4, there exists a substantial variation in teachers' educational background across countries. For example, some countries (France, Hungary, Netherlands, Spain) do not require teachers to complete college education while in other countries (Czech Republic, Finland, Slovak Republic, U.S.A.) more than half of math and science teachers hold M.A. or other professional degrees.

A more prominent discrepancy is observed in the proportion of teachers who are assigned to their main-field subjects, i.e., the subject in which he/she majored in college or in teacher training programs. Data indicate that

[12] Unfortunately, the first two waves of the TIMSS (1995 and 1999) do not provide detailed information on whether the teacher is state-certified.

Table 4
Various indicators for teacher's subject-matter knowledge.

	Sample size[a]	Math						Science						% of students: two subjects taught by the same teacher
		Education level		Main-field subject		Certification		Education level		Main-field subject		Certification		
		Have B.A. degree	Have M.A. degree	Main-field subject is math	Gender difference (male–female)	Have certification	Gender difference (male–female)	Have B.A. degree	Have M.A. degree	Main-field subject is science	Gender difference (male–female)	Have certification	Gender difference (male–female)	
Canada	8470	79.0%	14.6%	24.8%	1.4%	95.4%	5.9%	79.4%	14.8%	40.1%	7.6%	97.8%	3.7%	41.9%
Czech Republic	8001	0.2%	97.0%	60.6%	0	96.6%	−0.8%	0.6%	96.8%	82.9%	−0.8%	96.0%	−3.2%	30.9%
Finland	2957	37.8%	59.0%	80.1%	1.7%	92.2%	−4.9%	27.6%	69.1%	58.9%	3.0%	92.0%	−3.6%	37.3%
France	0	39.8%	26.2%	n.a.	n.a.	n.a.	n.a.	28.2%	49.8%	n.a.	n.a.	n.a.	n.a.	4.0%
Greece	0	98.4%	1.6%	n.a.	n.a.	n.a.	n.a.	95.0%	5.0%	n.a.	n.a.	n.a.	n.a.	0.0%
Hungary	10,371	51.2%	7.9%	79.3%	3.1%	100.0%	0.0%	56.2%	12.7%	88.8%	−4.0%	99.0%	0.3%	28.2%
Japan	15,171	93.9%	5.9%	81.3%	4.4%	100.0%	0.0%	88.8%	11.2%	83.7%	2.9%	100.0%	0.0%	0.0%
Netherlands	4919	53.5%	13.9%	66.6%	3.1%	97.8%	−1.7%	44.1%	32.9%	66.3%	−7.0%	94.3%	−5.2%	4.5%
Norway	8497	69.3%	8.0%	40.7%	0	97.9%	0.1%	69.4%	9.3%	51.3%	15.1%	97.9%	−0.3%	61.0%
New Zealand	7231	66.5%	15.6%	48.7%	14.2%	98.0%	−0.2%	62.6%	21.6%	81.8%	1.7%	98.8%	−2.0%	16.6%
Portugal	0	100.0%	0.0%	n.a.	n.a.	n.a.	n.a.	100.0%	0.0%	n.a.	n.a.	n.a.	n.a.	1.0%
Slovak Republic	7149	21.7%	77.5%	64.2%	8.2%	79.0%	14.1%	42.5%	56.7%	76.6%	8.8%	88.2%	3.8%	5.8%
Spain	0	67.5%	3.6%	n.a.	n.a.	n.a.	n.a.	68.1%	4.0%	n.a.	n.a.	n.a.	n.a.	58.4%
Sweden	9514	58.1%	22.6%	56.4%	0	91.2%	−7.3%	60.8%	26.6%	82.4%	2.0%	91.8%	1.3%	68.9%
U.S.A.	14,807	47.9%	52.1%	45.1%	−5.2%	97.2%	−3.9%	49.0%	51.0%	57.1%	12.3%	97.2%	−3.3%	6.3%

Notes: The teacher's education level, main-field subject and state-certification status information is from the TIMSS (1995, 1999, 2003, 2007). The main-field subject indicates the subject that the teacher majored in colleges. The teacher's main-field subjects and certification status are available in TIMSS 2003 and 2007.

[a] Sample size indicates the number of observations for which additional teacher controls (teacher's certification status and teacher's main-field subject) are available. Total sample sizes are reported in Table 1.

Table 5
The FE estimates with subject-matter knowledge controls.

	Boys			Girls		
	FE	FE		FE	FE	
		Math	Science		Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
Canada	0.02	0.00	0.05	0.01	0.01	0.00
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)
Czech Republic	0.00	0.01	−0.01	0.00	0.01	0.00
	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)
Finland	0.01	0.05	−0.01	0.01	0.01	0.02
	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)
France	0.00	0.01	0.00	0.03	0.07	0.00
	(0.02)	(0.04)	(0.02)	(0.03)	(0.05)	(0.02)
Hungary	0.02	0.02	0.01	0.00	0.00	0.00
	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)
Japan	0.01	0.01	0.02	0.01	0.01	0.02
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
Netherlands	−0.01	−0.02	−0.01	−0.01	0.00	−0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)
Norway	−0.03	−0.02	−0.04	0.00	−0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
New Zealand	0.02	0.03	0.03	0.00	−0.04	0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Portugal	0.01	−0.01	0.02	0.01	0.03	0.00
	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)
Slovak Republic	−0.01	−0.03	0.01	0.01	0.00	0.01
	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.02)
Spain	0.13	0.08	0.15	0.12	−0.01	0.07
	(0.11)	(0.07)	(0.10)	(0.10)	(0.07)	(0.09)
Sweden	−0.02	−0.02	−0.02	0.02	0.02	0.02
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
U.S.A.	0.00	0.00	0.00	0.00	0.01	−0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

Notes: Robust standard errors, adjusted for clustering on school level, are in parentheses. All models include teacher/classroom controls listed in the data section. Country-by-year dummies as well as country-by-grade dummies are included in all models. Test scores are standardized to an international mean of 0 and a standard deviation of 1. In addition, the models include indicators for teacher's main-field subject. In estimation, the students who are instructed math and science by a single teacher are excluded from the analysis. None of the estimates reported in this table is statistically significant.

teachers in Hungary or Japan have a stronger content background in the field in which they are teaching. Canadian and Norwegian teachers, on the other hand, are more likely to teach subjects out of their field of study. Such noticeable differences across countries may stem from variations in entry requirements for teacher education. Some countries have tightly configured and rigorously enforced standards that make it difficult for teachers to instruct subjects other than their field of study. However, in other countries, a teacher who majored in education is eligible to teach various science and math subjects. Furthermore, the proportion of teachers assigned to their main field subject differs between math and science. On average, science teachers are more likely to be assigned to their main field subject than math teachers. In Czech Republic, for instance, only 61 percent of math teachers are assigned to their main field subject, while approximately 83 percent of science teachers currently teach theirs.

A review of Tables 3 and 4 suggests that teacher's subject matter knowledge may partly account for the outcome of teacher effectiveness on student achievement. Table 3 indicates that in Sweden, for instance, a positive correlation between teacher gender and student achievement is statistically significant in science but not in math. Table 4 shows that, in Sweden, around 56 percent of math teachers majored in mathematics, while more than 82 percent of science teachers majored in science. In other words, it is possible that, in Sweden, science teachers are more likely to provide more effective teaching than math teachers. With no control for the difference in teacher's subject matter knowledge between the two subjects, the difference in teaching quality between math and science teachers may not be fully controlled for. Therefore, the observed positive correlation between female teachers and girls' test scores may be the result of differences in teacher quality.

5.2. The FE estimates with subject-matter knowledge controls

The importance of teacher subject-matter knowledge in positively influencing student achievement means that indicators for teacher's subject-matter knowledge level should be included in the specifications to fully control for teacher quality and efficacy. Among various indicators for teacher's subject matter knowledge, indicators for teacher's education level and experience are already included in the specifications. Unfortunately, the TIMSS did not begin to survey teacher's certification status and major subject field until the third wave (2003). Disappointingly, survey completion rates for these indicators vary significantly across countries. Consequently, the inclusion of the indicators for teacher's certification status and major subject field leads to a significant loss in observations. It is for this reason that the control variables for teacher quality were not fully controlled for in the results reported in Tables 2 and 3.

Despite the possibility of a loss of a substantial amount of observations, this section examines whether the inclusion of extensive teacher control variables explains the observed positive correlation between teacher gender and student achievement in a small number of countries. To further control for differences in teacher knowledge related to subject, this study employs a model with extensive controls for teacher quality. These controls include teacher certification, whether the teacher instructs his/her main field subject, and teacher's education level and years of experience. Additionally, students taught math and science by a single teacher are excluded from the analysis.[13] This is necessary as it suggests that the teacher may be teaching at least one out-of-field subject to students, thus the effectiveness of the teacher's instruction may differ between the two subjects.

The FE estimates for the gender matching effect with further teacher quality controls in sample countries are reported in Table 5. The effect of teacher's gender on boys' test scores in math and science are reported in columns 1 and 2, respectively, and the effect on girls' test in math and science are in columns 3 and 4, respectively. The results in Table 5 indicate that assignment to a same gender teacher does not significantly impact students' test scores in all sample countries except Greece.[14] The absence of teacher–student gender matching effect in these 14 countries confutes the hypothesis that assignment to a same gender teacher improves student's achievement, at least in math and science.

6. Conclusion

The debate concerning the relationship between gender gaps in students' academic achievement and teacher gender continues to disturb not only educators, but also society as a whole. Arguments blaming teachers for the existing gender gaps are seldom supported by empirical evidence from nationally representative samples. Furthermore, policies to increase the proportion of teachers of a certain gender have received intensified support over time, despite the mixed evidence regarding the consequences of such policies provided by empirical studies. The results of this study shed light on those policy issues and find no strong evidence to rationalize such policy initiatives. This paper provides information that may elucidate the outcomes of teacher–student gender interaction in the classroom. While a number of hypotheses suggest that teacher–student gender matching should have a positive impact on student performance, the results reported in this paper lend little credence to these hypotheses. First, the FE estimates indicate that, in eight of 15 OECD countries, teacher–student gender matching has no impact on students' test scores. The results of this study indicate that teacher gender is positively correlated with students' test scores at the 5 percent level in five countries and at the 10 percent level in two countries. Second, however, this paper presents evidence that the observed positive correlation between teacher gender and student achievement can be explained by the difference in teacher's subject matter knowledge level. Comprehensively, the empirical evidence presented in this paper provides no strong support for the hypothesis that teacher–student gender matching improves students' academic achievement.

This analysis, however, may provide a limited picture of teacher–student gender interaction. First, this paper examines the effect of teacher's gender in two subject areas in which gender gaps in test scores favor boys over girls. The outcomes of teacher–student gender interactions may differ in other subjects, for example, language subjects in which gender gaps favor girls. A further examination of the effect of teacher gender in reading and writing will provide considerable insight on the teacher–student gender interactions in the classroom. Second, this paper examines the effect of teacher's gender on test scores. However, test scores are only one component of various educational outcomes. Although little evidence was found to demonstrate that sharing the same-gender with teachers improves test scores, a same-gender teacher may positively influence other outcomes such as student's attitudes toward the subject (Dee, 2007), educational attainment decisions (Nixon & Robinson, 1999), course taking and choosing a major field at college (Rask & Bailey, 2002), and student behaviors. For future research, an extensive analysis of the effect of teacher–student gender matching on other educational outcomes is crucial to obtain additional information in the area of teacher–student interaction.

Appendix A.

Tables A1–A3.

[13] From Table 4 it becomes clear that, in some countries, students are instructed math and science by a single teacher. Take Norway, Spain, and Sweden for example, where, some teachers instruct math and science to their students in the same school year. If the teacher's main-field subject is mathematics, then it could be expected that the quality of teaching would be higher in math than in science.

[14] Greece only participated in the TIMSS in 1995 and information on teacher's main-field subject is unavailable. Consequently, the positive and significant FE estimate for Greek female students cannot be explained by the variables suggested above. Based on observations of the data, my theory on the Greece case is that 55 percent of students have one math teacher and two science teachers while the remaining students have only one math teacher and one science teacher. An examination of the data reveals that the positive correlation between teacher gender and student achievement for girls in math subjects is only observed among students with two science teachers. As a result, I speculate that there are certain differences in teacher/school traits between students with one science teacher and those with two. Notably, with the given data, any observable difference in teacher/school traits cannot be ascertained. Therefore, I leave the Greek case for further study with a more extensive data set.

Table A1
The FE estimates for other OECD countries.

	Boys		Girls	
	(1)	(2)	(3)	(4)
Australia	0.00	−0.02	0.00	−0.01
	(0.01)	(0.02)	(0.01)	(0.02)
Austria	0.03	0.04	−0.02	−0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Belgium	−0.02	−0.02	0.01	0.02
	(0.01)	(0.02)	(0.02)	(0.02)
Denmark	−0.06	−0.07	−0.04	−0.04
	(0.05)	(0.06)	(0.05)	(0.05)
England	0.01	0.01	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Germany	−0.03	−0.05	−0.01	−0.02
	(0.03)	(0.04)	(0.02)	(0.04)
Iceland	−0.05	−0.08	0.03	0.06
	(0.04)	(0.06)	(0.03)	(0.06)
Ireland	0.12***	0.05	−0.06**	−0.02
	(0.03)	(0.04)	(0.03)	(0.05)
Korea	−0.01	−0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Scotland	0.00	−0.01	0.01	−0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Slovenia	0.00	−0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Switzerland	0.00	0.05	−0.01	−0.04
	(0.04)	(0.05)	(0.04)	(0.05)
Additional teacher controls	X	O	X	O

Notes: Robust standard errors, adjusted for clustering on school level, are in parentheses. All models include teacher/classroom controls listed in the data section. In addition, the models include indicators for teacher's main-field subject. In estimation, the students who are instructed math and science by a single teacher are excluded from the analysis.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Table A2
Comparative review of the literature: math.

	Mathematics									
	Averages					Gender differences (male–female)				
	[NELS:88]	[TIMSS]				[NELS:88]	[TIMSS]			
	1988	1995	1999	2003	2007	1988	1995	1999	2003	2007
Standardized test scores										
Boys	0.04	−0.15	0.06	0.37	0.47	0.07***	0.08***	0.05**	0.08***	0.05**
	(1.02)	(1.01)	(0.91)	(0.91)	(0.92)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Girls	−0.03	−0.22	0.01	0.29	0.42					
	(0.97)	(0.96)	(0.87)	(0.86)	(0.90)					
OLS estimates										
Boys	0.07***	−0.03	0.07	0.16***	0.05					
	(0.03)	(0.07)	(0.06)	(0.06)	(0.06)					
Girls	−0.04	0.03	−0.03	−0.09	−0.04					
	(0.02)	(0.06)	(0.06)	(0.06)	(0.06)					
FE estimates										
Boys	−0.08***	−0.03	0.02	0.00	0.01					
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)					
Girls	−0.07**	0.01	0.03	0.00	0.02					
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)					
% of female teacher	57.6%	66.4%	59.3%	65.0%	66.9%					
Education										
B.A.	57.9%	57.0%	52.5%	38.7%	40.7%	−13%	0%	−1%	0%	4%
Graduate/professional degrees	40.9%	43.0%	47.5%	61.3%	59.3%	11%	0%	1%	0%	−4%
Years of experience										
1–3 years	11.1%	18.0%	11.8%	14.9%	15.6%	−3%	0%	−4%	0%	3%
4–6 years	9.3%	10.0%	10.1%	12.2%	12.9%	0%	5%	6%	2%	−5%

Table A2 (*Continued*)

	Mathematics									
	Averages					Gender differences (male–female)				
	[NELS:88]	[TIMSS]				[NELS:88]	[TIMSS]			
	1988	1995	1999	2003	2007	1988	1995	1999	2003	2007
7–9 years	10.0%	9.6%	11.3%	12.7%	13.7%	−6%	1%	−8%	−2%	−4%
10–12 years	11.2%	7.2%	8.8%	10.0%	12.6%	−7%	−1%	9%	1%	−1%
13–15 years	12.4%	7.3%	11.2%	11.1%	12.9%	−3%	−1%	−2%	0%	3%
16–18 years	14.0%	7.1%	6.9%	6.1%	5.5%	−4%	−6%	−3%	2%	3%
19–21 years	11.3%	11.9%	7.3%	5.8%	4.7%	3%	−5%	−2%	3%	0%
22–24 years	6.2%	8.9%	7.3%	5.8%	5.0%	4%	−3%	4%	−4%	0%
25 years or more	13.9%	20.1%	25.2%	21.4%	17.2%	15%	9%	1%	−2%	1%
% of teacher with certification	78.4%	–	–	97.6%	96.8%	1%	–	–	0%	0%
Main-field subject is math	42.3%	–	45.0%	47.3%	42.9%	−6%	–	−3%	−7%	−7%
Average class size	23.2	28.2	30.2	24.4	24.5	0.2	0.4	5.9	−0.8	0.6

Note: The data are from the National Educational Longitudinal Study of 1988 and the four waves of the Trends in International Mathematics and Science Study (1995, 1999, 2003, 2007).

Table A3
Comparative review of the literature: science.

	Science									
	Averages					Gender differences (male–female)				
	[NELS:88]	[TIMSS]				[NELS:88]	[TIMSS]			
	1988	1995	1999	2003	2007	1988	1995	1999	2003	2007
Standardized test scores										
Boys	0.10	0.15	0.28	0.60	0.52	0.17***	0.13***	0.17***	0.19***	0.14***
	(1.05)	(1.11)	(1.02)	(0.99)	(1.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Girls	−0.07	0.01	0.11	0.41	0.37					
	(0.93)	(1.02)	(0.92)	(0.92)	(0.94)					
OLS estimates										
Boys	0.06**	−0.07	0.07	0.05	−0.02					
	(0.03)	(0.07)	(0.05)	(0.06)	(0.06)					
Girls	0.00	0.08	−0.02	−0.03	−0.04					
	(0.03)	(0.06)	(0.05)	(0.05)	(0.05)					
FE estimates										
Boys	0.05*	0.02	0.07**	0.04	−0.02					
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)					
Girls	−0.02	0.03	0.02	−0.04	−0.01					
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)					
% of female teacher	48.7%	56.7%	50.8%	54.4%	56.3%	49%	57%	51%	54%	56%
Education										
B.A.	54.7%	58.1%	57.6%	40.1%	37.2%	−6%	2%	−8%	−3%	1%
Graduate/professional degrees	44.3%	41.9%	42.1%	59.9%	62.8%	6%	−2%	8%	3%	−1%
Years of experience										
1–3 years	13.3%	17.5%	16.0%	12.5%	14.6%	−1%	5%	−5%	1%	−4%
4–6 years	11.5%	15.1%	17.9%	16.1%	17.6%	−4%	−9%	−8%	−2%	3%
7–9 years	10.5%	9.2%	8.4%	16.7%	13.3%	−7%	−3%	−1%	−6%	1%
10–12 years	11.6%	10.4%	10.0%	11.0%	14.8%	−7%	−3%	−1%	0%	−5%
13–15 years	11.3%	6.4%	10.6%	5.0%	5.4%	−2%	−1%	−5%	1%	−4%
16–18 years	12.5%	7.7%	6.3%	5.6%	7.7%	3%	0%	1%	−3%	−3%
19–21 years	7.8%	9.0%	4.1%	6.5%	6.4%	7%	−4%	−4%	0%	−1%
22–24 years	6.7%	8.4%	3.6%	2.9%	5.0%	2%	−3%	1%	1%	3%
25 years or more	14.2%	16.4%	23.0%	23.6%	15.2%	9%	17%	21%	9%	10%
% of teacher with certification	79.6%	–	–	97.3%	97.0%	11%	–	–	−4%	−2%
Main-field subject is science	50.5%	–	52.4%	60.4%	58.8%	15%	–	18%	12%	7%
Average class size	24.0	37.4	28.7	25.1	28.4	0.0	–	0.2	0.1	0.1

Note: The data are from the National Educational Longitudinal Study of 1988 and the four waves of the Trends in International Mathematics and Science Study (1995, 1999, 2003, 2007).

References

American Association of University Women. (1992). *How schools shortchange girls*. Washington, DC: AAUW Education Foundation.

Almquist, E. M., & Angrist, S. S. (1971). Role model influences on college women's career aspirations. *Merrill-Palmer Quarterly*, *17*, 236–279.

Ammermuller, A., & Dolton, P. (2006). *Pupil–teacher gender interaction effects on scholastic outcomes in England and the USA*, ZEW Discussion Paper No. 06-060. Centre for European Economic Research.

Andersson, C., Johansson, P., & Waldenstrom, N. (2011). Do you want your child to have a certified teacher? *Economics of Education Review*, *30*(1), 65–78.

Ashenfelter, O., & Krueger, A. (1994). Estimating the returns to schooling using a new sample of twins. *American Economic Review*, *84*, 1157–1173.

Basow, S. A., & Howe, K. G. (1980). Role-model influence: Effects of sex and sex-role attitude in college students. *Psychology of Women Quarterly*, *4*, 558–572.

Becker, J. R. (1981). Differential teacher treatment of females and males in mathematics classes. *Journal of Research in Mathematics Education*, *12*, 40–53.

Begle, E. G., & Geeslin, W. E. (1972). *Teacher effectiveness in mathematics instruction, national longitudinal study of mathematical abilities reports*, No. 28. Washington, DC: Mathematical Association of America.

Braun, C. (1976). Teacher expectation: Socio-psychological dynamics. *Review of Educational Research*, *46*, 185–213.

Brophy, J. (1985). Interactions of male and female students with male and female teachers. In L. C. Wilkinson, & C. B. Marrett (Eds.), *Gender influences in classroom interaction* (pp. 115–142). FL: Academic Press.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*, 525–554.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the NELS 1988. *Industrial and Labor Relations Review*, *48*, 547–561.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher*, *18*, 4–10.

Goldhaber, D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*, 129–145.

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.

Irvine, J. J. (1986). Teacher–student interactions: Effects of student race, sex, and grade level. *Journal of Educational Psychology*, *78*, 14–21.

Kleinfeld, J. (1998). *The myth that schools shortchange girls: Social science in the service of deception*. Washington, DC: The Women's Freedom Network.

Krieg, J. M. (2005). Student gender and teacher gender: What is the impact on high stakes test scores? *Current Issues in Education*, *8*(9).

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review*, *28*(1), 49–57.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145.

Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review*, *27*(5), 669–689.

Nixon, L. A., & Robinson, M. D. (1999). The educational attainment of young women: Role model effects of female high school faculty. *Demography*, *36*, 185–194.

Ouazad, A. (2008). *Assessed by a teacher like me: Race, gender and subjective evaluations*, CEE Discussion Papers 0098. Centre for the Economics of Education, London School of Economics and Political Science.

Rask, K. N., & Bailey, E. M. (2002). Are faculty role models? Evidence from major choice in an undergraduate institution. *Journal of Economic Education*, *33*, 99–124.

Riordan, C. (1999). The silent gender gap. *Education Week*, *19*, 46.

Robst, J., Keil, J., & Russo, D. (1998). The effect of gender composition of faculty on student retention. *Economics of Education Review*, *17*, 429–439.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *Urban Review*, *3*, 16–20.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on student's achievement. *Sociology of Education*, *70*(4), 256–284.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247–256.

Sommers, C. H. (2000). *The war against boys: How misguided feminism is harming our young men*. New York: Simon and Schuster.

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