BUDAPEST WORKING PAPERS
ON THE LABOUR MARKET
BWP – 2017/2

Does teacher gender matter in Europe? Evidence from TIMSS data

ALFA DIALLO – ZOLTÁN HERMANN

BWP 2017/2

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, 2017

Budapest Working Papers on the Labour Market

BWP – 2017/2

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

Does teacher gender matter in Europe? Evidence from TIMSS data

Authors:

Zoltán Hermann
research fellow
Institute of Economics
Centre for Economic and Regional Studies, Hungarian Academy of Sciences
email: hermann.zoltan@krtk.mta.hu

Alfa Diallo
research associate
Regional Centre of Energy Policy Research
email: alfa.diallo@rekk.hu

January 2017

ISBN 978-615-5594-86-1
ISSN 1785 3788

Does teacher gender matter in Europe? Evidence from TIMSS data

Alfa Diallo – Zoltán Hermann

Abstract

This paper examines the effect of teacher gender on student achievement in 20 European countries. We employ a student fixed effect approach to account for unobservable subject-invariant student ability and non-random student-teacher sorting. Our results show that female teachers tend to increase students' test scores, especially for girls. However, this effect is far from universal; it is present in half of the countries in our sample. The female effect is likely to reflect selection into the teaching profession, as it is stronger in countries where the teacher wages relative to graduate wages are higher for women than for men. Having a teacher of the same gender also benefits students in Western Europe. We further find that the female teacher effect is more pronounced for low achievers, and in Western Europe for students with an immigrant background.

Keywords: teacher gender, student achievement, fixed-effect estimation, TIMSS

JEL codes: I21, J24

Acknowledgement:

This work was supported by the Hungarian National Research Development and Innovation Grant under Grant number K109338.

Számít a tanár neme Európában?
Eredmények a TIMSS adatok alapján

Alfa Diallo – Hermann Zoltán

Összefoglaló

A tanulmány a tanár nemének hatását vizsgálja a tanulói teljesítményekre 20 európai országban. Diák fix hatás modelleket alkalmazva vesszük figyelembe a nem megfigyelhető nem tantárgy-specifikus egyéni képességeket és a nem véletlenszerű tanár-diák párosítást. Az eredmények azt mutatják, hogy a női tanárok diákjai jobb teszteredményeket érnek el, különösen a lányok. Ez a hatás azonban nem univerzálisan jellemző, csak a mintabeli országok felében kimutatható. A becsült hatás feltehetően összefügg a tanári szelekcióval, ugyanis azokban az országokban erősebb, ahol a diplomások átlagbéréhez viszonyatott tanári bérek magasabbak a nők, mint a férfiak körében. A diákkal megegyező nemű tanárok hatása is pozitív Nyugat-Európában. Ugyanakkor a női tanárok hatása erősebb a gyengébb tanulók, és Nyugat Európában a bevándorló diákok esetében.

Tárgyszavak: tanár neme, tanulói teljesítmény, fix-hatás becslés, TIMSS

JEL kódok: I21, J24

Köszönetnyilvánítás: A kutatást az NKFIH támogatta, azonosító: K109338.

In the recent decades, girls have surpassed boys in education attainment and higher education enrolment in the developed countries (Pekkarinen, 2012). At the same time, they closed the test score gap in subjects traditionally favoring boys, like maths and science, while extending their advantage in reading literacy (Marks, 2008).

One possible factor behind girls' increasing success in school is the long-term trend of the growing share of female teachers, assuming that students benefit from having a teacher of the same gender. This explanation motivates the literature analyzing the effect of gender match on student achievement. Most of the literature focuses on the US. The results for having a same gender teacher in elementary and high school education vary from a generally positive effect (Nixon-Robinson, 1999; Dee, 2007) through mixed findings (Ammermueller-Dolton, 2006; Antecol et al, 2012, Winters et al., 2013) to no effect (Ehrenberg et al, 1995). Evidence for Europe is scarce and less supportive. While Ammermueller and Dolton (2006) found some positive same gender effects for the UK, Holmlund, and Sund (2008), Neugebauer et al. (2011) and Driessen (2007) reports no effect for Sweden, Germany, and the Netherlands respectively. Note that this literature focuses on the student teacher gender match and the overall difference in the effectiveness of female and male teachers receives little attention.

Another strand of related literature analyzing the effects of observable teacher characteristics on teacher effectiveness revolve around teacher education and experience and hardly investigates quality differences between male and female teachers (see the overview of Hanushek and Rivkin, 2006). Notable exceptions are Krieg (2005), Chudgar and Sankar (2008) and recently Escardíbul and Mora (2013), reporting a positive female teacher effect on student achievement.

This paper adds to the literature by providing new evidence on teacher gender effects in Europe. Our question is whether teacher gender matters for student achievement and if it does is there different effect for girls and boys. Our identification strategy relies on a student fixed effect approach; we compare test scores in different subjects taught by different teachers within students. This way the estimates are not affected by selection bias stemming from a non-random matching of students and teachers, as far as sorting is based on subject-invariant student ability.

We analyze the teacher gender effect in 20 European countries, using the TIMSS 2003, 2007 and 2011 datasets. This allows for the comparison of the effect across countries and relating the differences to other country characteristics.

Our contribution to the literature is twofold. First, we extend the student teacher gender match literature, mostly limited to the US and UK so far, with large-scale

European evidence. Second, we do not focus on the gender match exclusively, but interpret this in comparison with the general effect of having a female teacher. In other words, when teacher gender matters, what matters more, the gender match or the possibly different quality of female and male teachers?

Our main results suggest that teacher gender has no universal effect on student performance in Europe. First, it affects student achievement only in less than half of the countries in our sample. The effect is stronger and more frequently observed in Western Europe than in the East. In the former group of countries having a female teacher tends to raise student test scores. Second, when teacher gender matters the effect tends to be larger for girls; boys benefit less, if any, of having a female teacher. Third, there is a notable heterogeneity in the female teacher effect not only across countries but also across groups of students within countries. The female teacher effect is more pronounced for low achievers, and in Western Europe for students with an immigrant background. Finally, cross-country differences in the gender gap in the teacher wage penalty hint that the female teacher effect is related to selection into the teaching profession. In countries where teacher wages relative to average graduate wages are higher for women than for men, teacher gender is more likely to affect student achievement.

The remaining part of the paper is structured as follows. The next section provides an overview of the literature on the teacher gender effect. This is followed by a short description of the TIMSS dataset. The fourth section presents identification strategy and estimation method. Next, we present estimation results. This is followed by concluding remarks.

LITERATURE REVIEW

In the economic literature, many papers investigate the question, how the gender of the teacher affects students' performance in elementary and secondary school. The topic is quite complex as to draw firm conclusions one must consider the gender differences on both of the student and teacher side, as well as the sorting of pupils and teachers on their gender. In this section, we summarize the main theoretical considerations and the corresponding empirical evidence on these topics.

Our first research question is if students of male or female teachers perform better. Interestingly there are quite a few papers in which this basic question is in the main focus of the analysis while many papers tried to capture the effect of other observable characteristics of teachers (see the overview of Hanushek and Rivkin, 2006).

Eide et al. (2004) suggest a possible explanation of the potential performance difference between male and female teachers First, they show the portion of female teachers are much higher in elementary school, and secondary school than male ones. They also show that teachers are adversely selected, i.e. not the most talented persons choose to become teachers, and many of them leave their job seeking higher wages. This argument is widely accepted in the literature. Bacorod (2003), Dolton, (2006), Dolton-Marcenaro-Gutierrez, (2011), Varga (2017) also provides indirect evidence for adverse selection into the teacher profession.

If the relative rate of adverse selection is different between genders, this can result in performance differences between male and female teachers. Several mechanisms might explain a different rate of adverse selection. First, if in the case of men being a teacher is not a very prestigious job, the small portion of male teachers may mostly consist of men whose life goal is to teach, which will result in male teachers performing better on average. Another argument says that due to the unpopularity of teaching only the worst skilled men choose this profession, which results in better female teachers on average. Eide et al. (2004) present some evidence that men are relatively more adversely selected, than women as women have less alternative opportunity in the labor market than men. Finally, another possible explanation comes from Krieg (2005). According to his view, female teachers are better as they can encourage students more and can create a more positive atmosphere in the classroom.

There are only a few papers comparing the effectiveness of female and male teachers. The empirical evidence suggests that female teachers are more effective than men. Krieg (2005) shows this effect for eight years old students in mathematics, reading, and writing, Chudgar and Sankar (2008) on the sample 4th and 6th graders in language learning and Escardíbul and Mora (2013) also found evidence for 9th to 12th graders in mathematics. Muralidharan and Sheth (2013) also estimated a positive female teacher effect for primary school students in India using a value-added framework.

Interestingly the more specific question of the effect of the student teacher gender match on student performance is extensively researched. However, the evidence in this literature is mixed. Many authors find that there is a positive effect of same gender teachers on school performance (Dee, 2004; Ammermüller-Dolton, 2006; Dee, 2007), while there is another set of papers which cannot identify such positive pairing effect (Holmlund- Sund, 2008; Carrington et al., 2008, Neugebauer et al., 2011). Moreover, Antecol (2015) shows that in some cases it is possible, that same gender teachers have a negative effect on achievement.

Dee (2007) identifies two basic mechanisms that might explain a positive effect of same gender pairing. The first one affects the behavior of the students while the second is related to the attitudes of teachers. Dee's (2007) first theory is that student effort depends on whether the pupil considers his or her teacher as a role model. If that is the case the child is more willing to put more effort into school work, which results in better performance. Dee (2007) argues that it is more likely that a student chooses someone as a role model who is similar to him, e.g. is of the same gender. This mechanism is supported by the study of Bettinger and Long (2005), which provides evidence for this investigating university course selection.

The second mechanism is that teachers tend to discriminate students who are different (different gender) from them (Dee, 2007). This negative discrimination can hinder the performance of the pupil through two channels. First, directly as the student possibly cannot receive as much attention, and second, indirectly, if negative feedback makes the student invest less energy into learning. Greshenson et al. (2015) and Dee (2005) shows evidence that discrimination based on non-similarities exist. Others came to a different conclusion. Terrier's (2015) results suggest that teachers positively discriminate girls in grading in mathematics regardless of their gender, while Lavy (2004) presents results showing that at matriculation exam teachers discriminate against boys.

Finally, it is also possible, that the same gender effect is also driven by higher female teacher quality. If girls benefit more from teacher quality, female teachers may have a smaller positive effect on boys and a larger effect on girls. However, this mechanism is less likely in the light of the existing evidence on the effect of school quality. Autor et al. (2016) analyzed the effect of schools on the gender gap in Florida, and they found that boys benefit more from attending higher quality schools.

Altogether the debate is not yet settled, whether there is a positive benefit of same gender matching, and if so, what mechanism is primarily responsible for this effect.

DATA

Our analysis is based on data from the Trends in International Mathematics and Science Study (TIMSS). TIMSS is an international assessment program that was launched in 1995, and since then it is repeated in every four years. It measures the educational performance of 4th and 8th grade children in mathematics and science. TIMMS uses a two-stage sampling method. In the first stage, schools are selected from a given country based on a probability measure which is related to the number of students taught. In the second stage one or more classes are chosen from the previously

sampled schools, and in these classes all pupils participate in the tests. Each wave of TIMMS contains data from approximately twenty countries from all over the world, but the majority of the countries are East- and Central- European, Scandinavian, Asian and American countries. A shortcoming of the dataset is that participating countries differ significantly in the different waves.

Nevertheless, TIMSS has many advantages. First, TIMSS measures student achievement in a standardized framework. The tests and the assessment of students' responses are entirely independent of teachers in the given school. Hence grading bias can not influence test scores. Second, TIMMS makes it possible to link the individual pupils with their corresponding teacher of each subject measured. Third, data contain a large number of student, teacher and school background variables. Finally, in many countries 8th graders learn science in four separate subjects; physics, chemistry, biology and earth science, while student achievement is also measured in TIMSS separately for these fields. This provides ample variation of test scores and teacher characteristics within students.

A further advantage is that student performance is measured for quite similar subjects. As we have mentioned in the theoretical section, there is ample evidence that the pairing of students and teachers are not random (Kane et al., 2011). Therefore the identification strategy of most papers relies on the variation within student results to control for unobserved student ability. The more similar the subjects are, the less likely results are biased by non-random student-teacher matching based on subject specific ability.

We used 2003, 2007, 2011 waves of TIMSS. The control variables in 1995 and 1999 waves were quite different from that of the other waves, so we excluded these waves from our estimation sample.

We considered only European countries in the analysis, to compare countries with similar educational culture. In our sample we have distinguished; Western and Eastern European countries to determine whether the effect of teacher's gender is the same or not in these two regions of Europe.

Table 1 shows the list of countries, 8 Western European and 12 Eastern European, included in the analysis. There are important differences between the two regions. First, in the Eastern European region natural sciences are divided into more subjects while in Western Europe the same educational regime exists only in Belgium, Finland and in the Netherlands. In Norway, Spain, and Scotland natural science subjects are merged into integrated science. A mixed regime is in place in Sweden and England. In

the former most of the students study natural sciences in different subjects, in the latter integrated science is more prevalent.

In our evaluation, we only worked with the test scores of 8th grade students. Our student fixed-effect estimation strategy requires a fair amount of variation in teacher gender within students. However, the teacher profession is heavily dominated by women teachers in most of the countries at elementary school level. The share of women teachers is larger at lower grades than at upper grades. The 90% of teachers of the 4th graders in the sample is women in most of the countries which does not provide enough variance in the gender of the teachers to estimate reasonable effects. This ratio is only approximately 65% for 8th graders.

Table 1

The share of female teachers by subject

country	country, short name	maths	physics	Biology	chemistry	earth science	Integrated science
Western Europe							
Belgium, Flemish Community	bfl	0.750	0.651	0.688		0.688	
Basque Country, Spain	bsq	0.619					0.629
England	eng	0.529	0.181	0.862	0.573		0.522
Finland	fin	0.509	0.527	0.704	0.553	0.726	
Netherlands	nld	0.319	0.161	0.356	0.161	0.277	
Norway	nor	0.384					0.398
Scotland	sco	0.553					0.493
Sweden	swe	0.504	0.381	0.584	0.559		0.512
Eastern Europe							
Bulgaria	bgr	0.869	0.824	0.866	0.902	0.710	
Bosnia and Herzegovina	bih	0.572	0.564	0.645	0.702	0.552	
Czech Republic	cze	0.788	0.538	0.812	0.832	0.644	
Estonia	est	0.896	0.575	0.892	0.866	0.851	
Hungary	hun	0.826	0.683	0.773	0.843	0.655	
Lithuania	ltu	0.932	0.685	0.916	0.936	0.806	
Latvia	lva	0.920	0.634	0.938	0.896		
Macedonia	mkd	0.641	0.625	0.630	0.665	0.454	
Romania	rou	0.571	0.717	0.830	0.838	0.622	
Serbia	srb	0.596	0.622	0.788	0.779	0.656	
Slovak Republic	svk	0.787	0.693	0.851	0.829	0.781	
Slovenia	svn	0.823	0.574	0.908	0.921	0.773	

Note: The share of female teachers is calculated weighted by the number of students taught.

Table 1 also summarizes the share of female teachers by countries and by subjects. In Western Europe the share of male teachers is generally higher, on average almost half of the teachers are men. In the Netherlands and Norway male teachers form the majority (75% and 61% respectively). In Eastern Europe the average share of females are 75% percent, so the profession is heavily dominated by women. The ratio of females is highest in Lithuania and Latvia with a corresponding value of approximately 85%.

The share of female teachers is very different by subjects as well in the two regions. In Eastern Europe biology and chemistry are relatively more women dominated, in physics and earth science the share of male teachers is much higher, while there is no difference in mathematics. On the other hand, the male-female ratio in the Western European countries are more balanced. In those education systems where natural sciences are merged into integrated science the number of male and female teachers are very similar. In those western countries where natural science is split into different subjects, we were not able to identify any general pattern.

Additionally, our student fixed-effect estimation strategy requires a fair amount of variation in teacher gender within students. Calculating the number and share of students having both female and male teachers indicates that this variation is substantial (Table A1). The share typically ranges between the values of 40% and 70%. The biggest variation is associated with Bosnia with 85% percent while it is smallest in Norway, 19%.

It is important to note that there is no variation in a non-negligible subset of the sample. Due to the school- and class –level sampling the observed students cluster together. As the share of female teachers may correlate with school characteristics (e.g. church or public, rural or urban, etc.), the samples we effectively use to identify the teacher gender effect might not perfectly represent the full student and teacher population.

Moreover, teacher gender is not observed in some cases, although the share of teachers with missing data is moderate. For most countries, less than 2% of the students were dropped due to missing teacher data. A higher share of students (4.5-6%) was excluded in England, Scotland and Sweden. Table A1 shows the sample size of students and teachers in the final estimation sample by country. Note that sample sizes are fairly large, but the number of primary sampling units, classes is moderate.

In order to compare different subjects to each other we standardized the tests scores within countries and waves, so in each country in each wave for each subject the test scores have 0 mean and a standard deviation of 1. Table A2 and A3 reports the distribution of test scores by teacher's and student's gender. In the Western European

countries students of male teachers seem to perform similarly to females in mathematics, and male teachers tend to dominate natural sciences in most of the countries. In the Eastern European region however female teachers tend to have an advantage in both mathematics and science. Regarding student gender, in both regions boys seem to achieve better results in science and girls in mathematics.

IDENTIFICATION STRATEGY

The fundamental methodological problem in estimating teacher gender effects, and teacher effects in general, is the non-random pairing of students and teachers (Clotfelter et al., 2006, Kane et al.,2011). Sorting based on student quality may occur both between and within schools. As far as high ability students are matched to female (or male) teachers more often than low achievers, the estimated effect of teacher gender on student achievement in an OLS specification will be biased.

There are two possible cases where the non-random sorting can be a methodological problem. The first one is if the quality of teachers is not independent of gender. In this scenario female teachers are better for some reason, and because of this, they are paired with more talented students.

The second possible case is if the quality of teachers is the same across genders, but the sorting for some reason is not independent of it. For example, let us assume that due to historical reasons school principals expect male teachers to be better. In this case, more prestigious schools which can choose from many applicants for teaching positions are more willing to hire men than women based on this belief. At the same time, these popular schools also enroll students with a more favorable family background. In this case, there is no difference between the two genders, but as a consequence of selection a standard estimation will result in a negative female teacher effect.

A common solution to the sorting bias is to identify the effect of the teacher by comparing the results of the same student across subjects. This methodology was used by Ammermüller and Dolton (2006) who investigated the gender difference of 4th and 8th graders in the United States and Great- Britain based on a comparison of mathematics and science results and Dee (2007) who compared the test results of 8th graders in the subjects of mathematics, science, reading and social science in the US.

We employ the same identification strategy, comparing test scores across subjects within students. Within-student estimation amounts to controlling for subject-invariant student ability and other student characteristics that affect performance in every subject similarly.

It has to be noted that one potential source of bias remains. It is not inconceivable that students are sorted into classes or study groups by subject-specific skills and better ability or more motivated groups are matched more frequently to either male or female teachers. We consider this type of sorting highly unlikely to seriously bias our results for two reasons. First, math and the science subjects are more akin to each other than for example math and reading, a comparison often used in the literature. Second, selection based on subject-specific skills is most likely in advanced level groups. We observe the actual number of lessons per week in each subject, and this allows us to control for groups studying at an advanced level directly. We identify advanced level groups as having at least 25% more lessons per week than the median weekly hours in the given subject in the country.

To explore the female teacher effect we first estimate a baseline OLS model of student achievement:

$$(1)\ y_{i,k} = \alpha + \beta_1 FT_{i,k} + \beta_2 FS_i + \gamma' TC_{i,k} + \sigma' SC_i + \varphi' GC_{i,k} + \lambda_k + \kappa_k + \epsilon_{i,k}$$

where $y_{i,k}$ denotes test score in subject k for student i, $FT_{i,k}$ is dummy variable taking the value 1 if the teacher in subject k is female, FS_i is an indicator variable for female students, SC_i, $TC_{i,k}$ and $GC_{i,k}$ are sets of student, teacher and study group characteristics respectively, λ_k is a set of indicator variables for each subject, κ_k is a set of indicator variables for each subject in case of female student, and $\epsilon_{i,k}$ is a normally distributed iid error term.

The student control variables are the education level of the student's father and mother, the number of books at home, and whether the student or the parents of the student were born in another country. All of these factors are measured as categorical variables. The teacher control variables are age, teaching experience, the level of education relative to the median teacher in the country and whether the teacher is qualified for teaching the given subject. Finally, the group level control variables are the number of students in the group, the number of lessons in a week and whether the group is an advanced level group.

Moreover we also control for subject fixed effects and their interactions with the female student indicator. Allowing for the subject fixed effects to differ by gender is especially important, since girls and boys tend to excel in different sciences. Moreover, as Holmlund and Sund (2008) suggest female teachers might be overrepresented in subjects preferred by girls. Hence, omitting subject-student gender interactions could result in estimating a spurious female teacher effect.

Our preferred specification, including student fixed-effects, is:

$$(2)\quad y_{i,k} = \alpha + \beta_1 FT_{i,k} + \gamma' TC_{i,k} + \varphi' GC_{i,k} + \eta_i + \lambda_k + \kappa_k + \epsilon_{i,k}$$

where η_i captures student fixed effects. We first estimate this model for the full sample of students to measure the overall effect of having a female teacher.

Then we estimate the same model separately by student gender, omitting κ_k , in order to detect potentially different female teacher effects on girls and boys. Note that a larger effect on girls than on boys indicates that a positive gender match of the teacher and the students improves student achievement.

Finally, we test the effect of having a teacher of the same gender by testing formally the difference in the female teacher effect on girls and boys:

$$(3)\ y_{i,k} = \alpha + \beta_1 FT_{i,k} + \beta_3 FT_i * FS_i + \gamma' TC_{i,k} + \varphi' GC_{i,k} + \pi' TC_{i,k} * FS_i + \theta' GC_{i,k} * FS_i + \eta_i + \lambda_k + \kappa_k + \epsilon_{i,k}$$

where β_3, the coefficient of the female student – female teacher interaction term is of key interest. Note that in order to preserve the full flexibility of estimating the fixed effect model separately by gender, a full set of interaction variables of teacher and group characteristics with the female student indicator are also included in the model.

We first estimate each of the above-presented equations for the pooled samples of Western European and Eastern European countries. These estimates are intended to describe the average teacher gender effects in the two groups and reveal the differences in the effects. Moreover, the pooled samples are large enough to precisely estimate even small effects.

Next, we repeat the estimation for each country separately in order to explore whether teacher gender effects are similar in most European countries. Note that these estimates are less precise due to the smaller number of classes per country.

All of our estimates are weighted. We rescaled the TIMSS student weights both within and across countries. Within countries the sum of rescaled weights is equal for each wave of the survey. At the same time, the sum of rescaled weights is also equal for each country, i.e. countries are equally weighted in the pooled estimates. Note that equal weighting is independent of how many waves a country took part in the survey.

In other words, countries with more survey years have weighted share in the pooled sample as the countries observed only once.

Finally, each regression is estimated with robust standard errors clustered at the class level. This way we take into account that the error terms may be correlated within the primary sampling units.

RESULTS

In the section, we present first the main results of our estimated regressions. Table 2 provides the estimates for the pooled samples of Western and Eastern European countries, while Table 3 and 4 lists the results estimated separately for each western and eastern country respectively.

Each table presents only the key coefficients of our three specifications, and each cell corresponds to separate regression estimate. Column 1 contains the baseline female teacher coefficient of the OLS specification (eq. 1), column 2 displays the female teacher coefficient of the student fixed-effect specification (eq. 2), columns 3 and 4 shows the same coefficient estimated separately for the sample of boys and girls, and column 5 includes the coefficient of the interaction term in eq. 3, which is a formal test of the equality of the coefficients for girls and boys.

Simple OLS estimates suggest that the students of female teachers perform better than the students of male teachers in Eastern Europe (Table 1, column 1). The estimated effect is similar for the group of Western European countries, but it is estimated very imprecisely and statistically not significant. However, OLS estimates are prone to be affected by sorting bias.

Our student fixed effect estimation strategy accounts for subject-invariant unobserved student ability and non-random student-teacher matching based on this. Overall, the results are not very far from the OLS estimates (column 2). In Western Europe the female teacher coefficient is slightly smaller but statistically significant. On the other hand in Eastern Europe the female teacher effect is also statistically significant but amounts to less than half of the OLS estimate. These patterns suggest that the non-random pairing of teachers and students related to gender is much more relevant in Eastern Europe.

Table 2

The estimated female teacher effect on test scores, pooled samples

country	OLS	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)	(5)
Western Europe					
	0.0194	0.0158**	0.000899	0.0305***	0.0296***
	(0.0158)	(0.00620)	(0.00748)	(0.00774)	(0.00900)
Eastern Europe					
	0.0247**	0.00939***	0.00472	0.0143***	0.00958
	(0.0125)	(0.00364)	(0.00437)	(0.00497)	(0.00587)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(4): coefficient of having a female teacher. Specification (5): coefficient of the interaction between female student and having a female teacher. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Altogether, the effect size is similar in the Western and Eastern European samples; female teachers improve student achievement by 0.016 and 0.009 standard deviations (SD).

To answer the question whether same gender pairing of students and teacher have any additional effect, we have to compare the estimates for boys and girls (column 3 and 4). In both country groups the female teacher effect is statistically significant only for girls. In Western Europe girls score 0.03 SD higher in subjects taught by a female teacher, while in the East the effect size about half of that.

Column 5 reports the coefficient of the female teacher – female student interaction term which enables us to formally test whether the female teacher effect is different for boys and girls. The interaction term is statistically significant for the West, indicating that having a teacher of the same gender has a positive effect on student achievement. At the same time in Eastern Europe the female teacher effects for boys and girls do not differ statistically significantly.

Altogether female teachers tend to improve test scores in both regions on average, especially for girls. In Western Europe the student teacher gender match also appears matter, while in the East we are unable to detect a significant effect.

In order to explore whether the female teacher effects are a common feature of most European education systems, we estimated the same models for country separately. Table 3 and 4 presents the coefficients for Western and Eastern countries respectively.

In Western Europe we can detect a significant teacher gender effect in half of the countries. In the Flemish Community of Belgium, the Basque Country and Norway

female teachers improve student achievement for either the full sample or the boy or girl subsample. In England having a female teacher goes together with lower test scores for boys, while girls seem to be not affected. Moreover, England is the only country in the Western group where the teacher gender effects statistically significantly differ by the gender of the student. Thus in England female teachers do not appear to teach more effectively in general, only the same gender effect is present.

However, it is important to note that due to the relatively small sample sizes regarding classes the estimates are imprecise and this may contribute to the lack of significant effects. For some countries the coefficients are comparable in magnitude to those estimated for the pooled sample, but accompanied by large standard errors (see e.g. the Netherlands). Moreover, though the interaction terms are not significant in any country that England, the same gender effect estimated for the pooled Western European sample is not driven solely by the English sample. Re-estimating the models for the pooled sample excluding England still provides a significantly stronger effect for girls[1]. Thus the same gender effect is probably relevant in several Western European countries.

The results are similarly mixed for Eastern Europe. Female teachers have some significant positive effect in half of the countries; Bulgaria, Hungary, Macedonia, Romania, Serbia, and Slovakia. The effect tends to be significant for girls, but not for boys, except Bulgaria. However, the same gender effect is significant only for Hungary and Slovakia. In the other countries the point estimates are very low, so this effect seems to occur rather as an exception in the Eastern European region.

Finally, it has to be noted, that multiple inference requires caution when interpreting statistical significance of the coefficients at face value. When the same regression model is estimated on many subsamples, some significant results can be expected to occur as a statistical artifact due to random error (Bland and Altman, 1995). However, the share of countries with significant coefficients is large enough in both Western and Eastern Europe to consider these false positive.

Our main results revealed substantial differences in the teacher gender effects across countries. With the next set of estimates we explore the heterogeneity of the teacher gender effect within countries, with respect to three student characteristics. First, we compare low and high achievers, then we turn to subgroups by family background and immigrant status.

[1] Results are available from the authors upon request.

Table 3

The estimated female teacher effect on test scores, Western Europe

country	OLS	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)	(5)
bfl					
	-0.0533	0.0405***	0.0214	0.0577***	0.0363
	(0.0460)	(0.0154)	(0.0185)	(0.0195)	(0.0237)
bsq					
	-0.00870	0.0508*	0.0431	0.0574	0.0143
	(0.0450)	(0.0271)	(0.0320)	(0.0351)	(0.0400)
eng					
	0.0259	-0.00734	-0.0241**	0.00920	0.0333**
	(0.0385)	(0.00881)	(0.0112)	(0.0109)	(0.0137)
fin					
	0.0124	0.00153	-0.00544	0.0111	0.0166
	(0.0341)	(0.0110)	(0.0141)	(0.0127)	(0.0165)
nld					
	-0.000338	0.0129	-0.00133	0.0298	0.0311
	(0.0752)	(0.0211)	(0.0233)	(0.0234)	(0.0208)
nor					
	0.0400	0.0371**	0.0443*	0.0308	-0.0135
	(0.0279)	(0.0183)	(0.0241)	(0.0234)	(0.0305)
sco					
	0.0165	-0.00136	0.00142	-0.00535	-0.00677
	(0.0406)	(0.0142)	(0.0190)	(0.0175)	(0.0233)
swe					
	0.00497	0.00693	-0.000134	0.0144	0.0145
	(0.0231)	(0.0106)	(0.0131)	(0.0126)	(0.0147)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(4): coefficient of having a female teacher. Specification (5): coefficient of the interaction between female student and having a female teacher. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 4

The estimated female teacher effect on test scores, Eastern Europe

	OLS	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)	(5)
bgr					
	-0.0994*	0.0220	0.0370*	0.00446	-0.0325
	(0.0566)	(0.0184)	(0.0196)	(0.0241)	(0.0229)
bih					
	0.0869**	-0.00318	-0.0173	0.0119	0.0292
	(0.0388)	(0.0139)	(0.0167)	(0.0168)	(0.0187)
cze					
	-0.0652*	-0.00298	-0.00114	-0.00466	-0.00352
	(0.0385)	(0.0112)	(0.0116)	(0.0164)	(0.0172)
est					
	-0.0147	0.0297	0.0260	0.0316	0.00560
	(0.0474)	(0.0197)	(0.0211)	(0.0269)	(0.0287)
hun					
	0.0376*	0.00988	0.00191	0.0181***	0.0162*
	(0.0196)	(0.00629)	(0.00880)	(0.00701)	(0.00980)
ltu					
	-0.0270	0.0108	0.00842	0.0138	0.00535
	(0.0253)	(0.00903)	(0.0109)	(0.0124)	(0.0149)
lva					
	0.0164	-0.0119	-0.00624	-0.0172	-0.0110
	(0.0559)	(0.0187)	(0.0254)	(0.0218)	(0.0292)
mkd					
	0.257***	0.0233**	0.0180	0.0293**	0.0113
	(0.0364)	(0.0106)	(0.0127)	(0.0123)	(0.0133)
rou					
	-0.0232	0.00845	-0.00103	0.0181*	0.0192
	(0.0260)	(0.00829)	(0.00990)	(0.0105)	(0.0119)
srb					
	0.0973***	0.0206**	0.0141	0.0270**	0.0129
	(0.0234)	(0.00812)	(0.00962)	(0.0108)	(0.0123)
svk					
	0.0614	0.00948	-0.0172	0.0382*	0.0554**
	(0.0463)	(0.0182)	(0.0213)	(0.0225)	(0.0247)
svn					
	0.0842**	-0.00275	0.00826	-0.0148	-0.0230
	(0.0370)	(0.0127)	(0.0154)	(0.0154)	(0.0175)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(4): coefficient of having a female teacher. Specification (5): coefficient of the interaction between female student and having a female teacher. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

First, we estimate teacher gender effects for low and high achiever students. We calculate mean test scores across subjects for every student and define low achievement as having a mean score below or equal to the median value in the country. Though this measure captures both student ability, motivation and teacher effects, we assume that it correlates with ability strongly. Table 5 presents estimation results for the pooled samples, while results for each country are shown in the Appendix (Table A4 and A5). In Western Europe the female teacher effect is clearly stronger for the low achievers among girls; the effect size is three times of that estimated for high achievers. The coefficients for boys are statistically not significant in either case, and the student-teacher gender match has a significant positive effect only for low achievers. The pattern of results is similar in Eastern Europe, though the differences between the coefficients for low and high achievers are much smaller and statistically not significant.

In Western Europe estimates by country mostly echo the results for the pooled sample (Appendix, Table A4). Teacher gender matters more among low achievers in Belgium, the Basque Country, and Norway. The only exception is England, where the negative female teacher effect on boys is stronger in the high achiever group. In Eastern Europe the results are more diverse across countries (Appendix, Table A5). The teacher gender effect is larger for low achievers in some cases (Estonia, Bosnia and Herzegovina), while the opposite pattern occurs in other countries (Bulgaria, Macedonia), and the effect is independent of the performance level of students in some countries (Hungary, Serbia).

Next, we estimate the female teacher effect for subgroups by family background, measured by the number of books at home. Students with more books than the median value in the country are considered to have a more favorable family background. Results for the pooled samples do not differ markedly by family background (Table 6). The difference in the effect size between boys and girls appears to be slightly larger among students with a less favorable family background, but the difference of the interaction terms is statistically not significant.

However, family background does seem to interact with the teacher gender effect in some countries (see Table A6 and A7 in the Appendix). Female teachers tend to improve test scores among student from more disadvantaged families in the Basque country, Norway, and Serbia. At the same time, female teachers affect students from better-off families in Scotland, Sweden, Bulgaria, Estonia, and Romania. In Slovenia, female teachers impair the scores of students with a more favorable background.

Finally, we compare students with an immigrant background with the native population. Immigrant background refers to students whom themselves or one of their

parents was born in another country. In Western Europe immigrant background amplifies the teacher effect markedly (Table 7). Female teachers improve the test scores of girls with immigrant background about two times compared to native girls. Since the performance of boys is not related to teacher gender in either group, the gender match between the student and the teacher seems to be more important among immigrant students. At the same time, it is important to note that the estimates for the full population are not driven by immigrant students; the estimates in Table 2 are very similar to the results for non-immigrant students.

Table 5

The estimated female teacher effect on test scores for low achievers and high achievers, pooled samples

	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)
Western Europe				
low achievers				
	0.0212**	-0.00513	0.0463***	0.0514***
	(0.00929)	(0.0115)	(0.0115)	(0.0142)
high achievers				
	0.0115*	0.00744	0.0153*	0.00782
	(0.00692)	(0.00859)	(0.00864)	(0.0103)
Eastern Europe				
low achievers				
	0.0114**	0.00404	0.0189***	0.0149*
	(0.00457)	(0.00560)	(0.00639)	(0.00779)
high achievers				
	0.00734	0.00572	0.00930	0.00357
	(0.00469)	(0.00602)	(0.00615)	(0.00777)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(3): coefficient of having a female teacher. Specification (4): coefficient of the interaction between female student and having a female teacher. Low/high achievers are students with average test scores over all subjects below/above the median student of their country. Robust standard errors clustered at the class level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6

The estimated female teacher effect on test scores for students with few and many books at home, pooled samples

	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)
Western Europe				
few books				
	0.0156**	-0.000803	0.0333***	0.0341***
	(0.00739)	(0.00911)	(0.00957)	(0.0117)
many books				
	0.0177**	0.00673	0.0271***	0.0204*
	(0.00757)	(0.00981)	(0.00965)	(0.0123)
Eastern Europe				
few books				
	0.00732*	0.00162	0.0136**	0.0120*
	(0.00422)	(0.00501)	(0.00606)	(0.00718)
many books				
	0.0121**	0.00809	0.0158**	0.00774
	(0.00503)	(0.00673)	(0.00674)	(0.00895)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(3): coefficient of having a female teacher. Specification (4): coefficient of the interaction between female student and having a female teacher. Few/many books refers to the number of books at home below or equal to / above the median value of the country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Conversely, in Eastern Europe the teacher gender effect is independent of immigrant background. Note, however, that immigrant background has a very different meaning in Western and Eastern Europe. In the latter group the share of immigrant students is larger in countries that gained independence after 1990; the Baltic countries and the parts of the former Yugoslavia. In these countries a parent born outside the country implies less difference, if any, compared to the native population in economic, cultural or linguistic terms than immigrant status in Western Europe.

Table A8 and A9 in the Appendix presents estimates by country for countries with a high share (above 15%) of students with an immigrant background. Due to the relatively small sample sizes estimates for the immigrant subsamples are imprecise and hardly significant. A notable exception is Norway, where female teachers improve the test scores of immigrant girls substantially, while immigrant boys are not affected. At the same time in the native sample only boys benefit from having a female teacher.

Altogether the teacher gender effect shows remarkable heterogeneity with respect to student characteristics. Female teachers have a stronger positive impact on low achievers and in Western Europe on students with an immigrant background.

Overall our estimation results show a mixed picture, indicating that the female teacher effect is far from universal; it is present in some countries, and not detectable in others. This implies the question why female teachers are more effective in some countries. One possible explanation for these cross-country differences is that selection into the teaching profession is different among men and women in some countries. We can not test this explanation directly, but indirect evidence is in line with this argument.

Table 7

The estimated female teacher effect on test scores for students with immigrant and non-immigrant background, pooled samples

	FE full	FE boys	FE girls	FE interaction
	(1)	(2)	(3)	(4)
Western Europe				
immigrant				
	0.0290**	0.00511	0.0542***	0.0491**
	(0.0133)	(0.0169)	(0.0160)	(0.0197)
non-immigrant				
	0.0133**	0.000445	0.0259***	0.0255***
	(0.00635)	(0.00764)	(0.00811)	(0.00943)
Eastern Europe				
immigrant				
	0.0144*	0.0122	0.0169	0.00463
	(0.00861)	(0.0108)	(0.0123)	(0.0154)
non-immigrant				
	0.00850**	0.00335	0.0139***	0.0106*
	(0.00375)	(0.00470)	(0.00500)	(0.00618)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate. Specification (1)-(3): coefficient of having a female teacher. Specification (4): coefficient of the interaction between female student and having a female teacher. Immigrant background refers to either student or mother or father born in another country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

The existing evidence suggests that teacher wages are a major factor in selection into teaching (Dolton, 2006). Several papers have found a positive association between teacher wages and student achievement in a cross-country setting, as well (Dolton-Marcenaro-Gutierrez, 2011; Varga, 2017). The authors suggest that the main mechanism driving these results is selection: higher wages attract more able teachers.

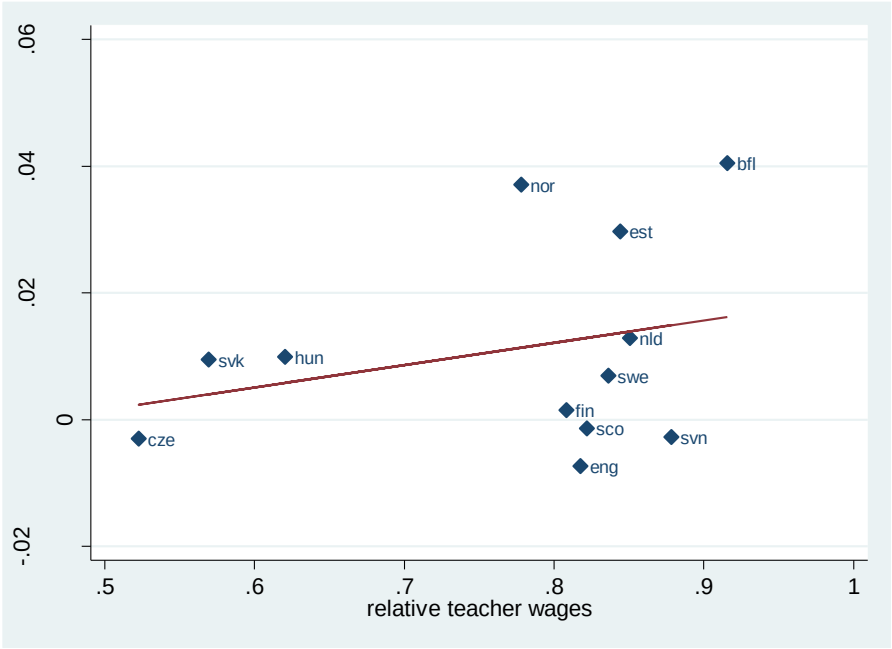
Looking for indirect evidence on we explored whether the level of teacher wages are related to the magnitude of the female teacher effect. We use average actual teacher salaries data provided by the OECD (OECD, 2015). Though these are available only for a few countries, we prefer actual salary data for several reasons. First, as Varga (2017) shows, actual salaries diverge substantially from the salary scale ratings in some countries. Second, OECD (2015) reports actual teacher salaries relative to the wages of workers with similar educational attainment. Finally, actual relative salaries are provided separately by gender, as well. Note that though the salary scale is identical for female and male teachers, actual relative teacher salaries may differ due to both different supplementary payments and seniority position of female and male teachers and the gender wage gap in the graduate labor market in general.

Figure 1 contrasts the overall relative teacher salary indicator to the female teacher effect, estimated on the full sample (top panel) and for girls only (bottom panel). Relative teacher wages are calculated here for the entire teacher population, men and women together. The female teacher effect estimated for the full sample seems to be slightly smaller when relative teacher wages are lower. However, this association is very weak and is driven by the huge difference in teacher wages between eastern and western European countries. In Slovakia, the Czech Republic and Hungary average teachers salaries amount to 50-60% of the wages of workers with similar educational attainment. In Western Europe, and also in Estonia and Slovenia, relative teacher wages fall into the range between 80-90%. Regarding the female teacher effect estimated on the sample of girls, there is no apparent relationship with relative teacher wages (Figure 1, bottom panel).

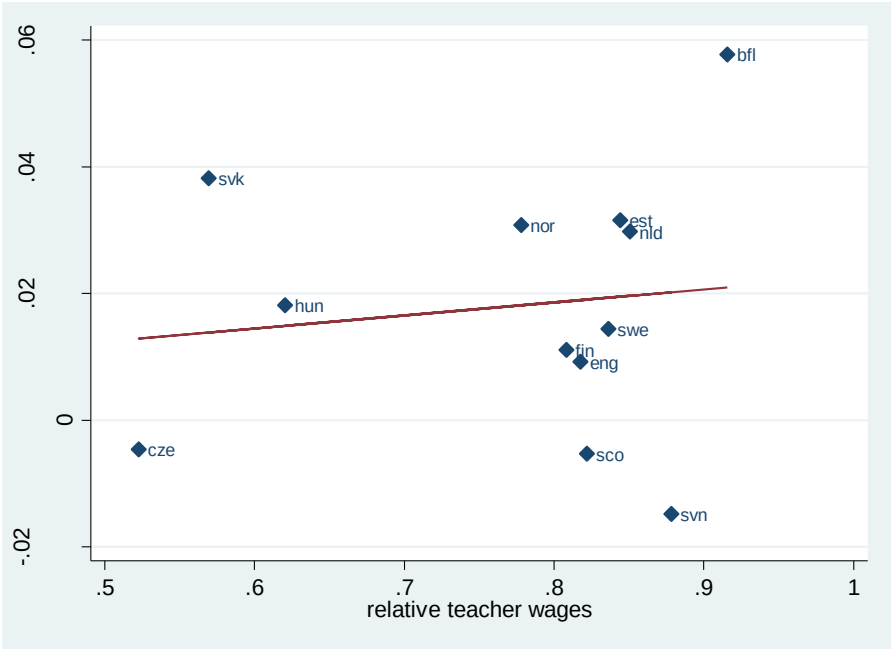
Figure 1

The estimated female teacher effect and relative teacher wages

Female teacher effect, full sample



Female teacher effect, girls



Notes: Estimated female teacher effects from Table 2 and 3. Relative teacher wages denote the ratio of annual average salaries of teachers at the lower-secondary level in public institutions relative to the wages of workers with similar educational attainment in 2013. Source: OECD (2015).

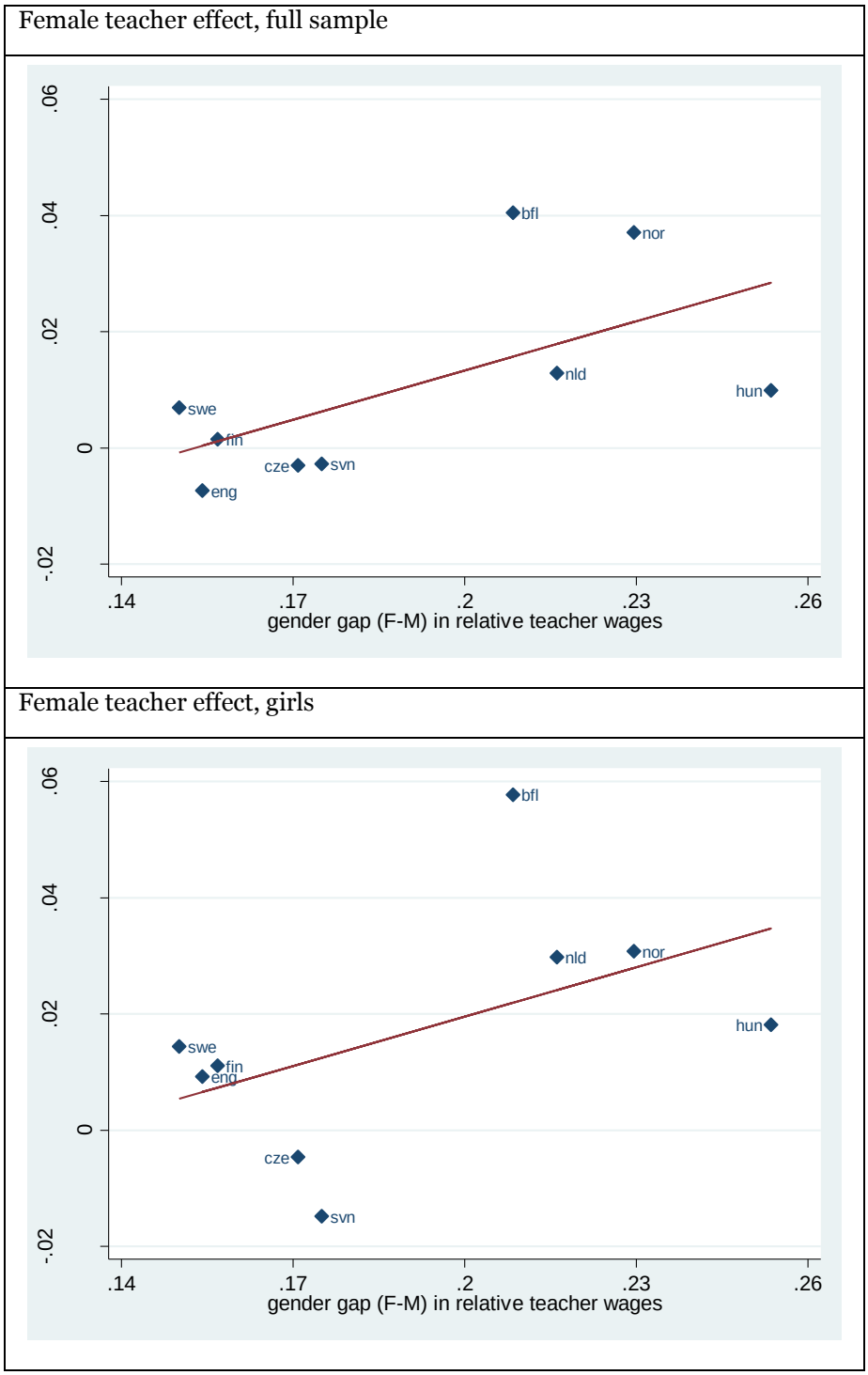
Altogether, the female teacher effect seems to be not related to the level of relative teacher wages. However, relative teacher wages do differ between men and women, and the gender gap varies over countries. We calculate the teacher gender wage gap as the difference of relative teacher wages between female and male teachers. Figure 2 depicts the female teacher effect and the gender wage gap. Regarding either the female teacher effect estimated on the full sample or the sample of girls, two groups of the countries stand out. In Belgium (Flemish Community), Hungary, the Netherlands, and Norway the gender gap is relatively large. Women lose about 20 percentage point less than men by choosing the teaching profession. In fact, in Belgium and the Netherlands female teachers earns 3-4% higher wages than the average women in a graduate job, while male teachers face a substantial wage penalty (17 and 19%, respectively). At the same time, in these four countries the female teacher effect is relatively large, and for Belgium, Norway and Hungary it is statistically significant for either the full sample or the sample of girls.

In the other group of countries the relative wages of female and male teachers are more similar, while the female teacher effect is small, and statistically not significant. Altogether a larger teacher gender wage gap seems to go together with female teachers more effective than men. Moreover, this relationship is not driven by a difference between eastern and western countries.

Since the actual teacher wage data are available only for a handful of countries, and these data are calculated for a single year, the association with the female teacher effect can not be considered more as an illustration. However, these stylized facts are in line with the selection hypothesis and suggest that female teachers are indeed more effective than men where it is worth to choose teaching more for women.

Figure 2

The estimated female teacher effect and the gender gap (F-M) in relative teacher wages



Notes: Estimated female teacher effects from Table 2 and 3. Relative teacher wages denote the ratio of annual average salaries of teachers at the lower-secondary level in public institutions relative to the wages of workers with similar educational attainment in 2013. Source: OECD (2015).

CONCLUSIONS

In this paper we investigated the effect of teacher gender on student performance in several European countries using the TIMMS international dataset. We employed a student fixed effect estimation strategy to account for unobserved subject-invariant student ability and non-random student - teacher sorting.

Our main result is that both in Western and Eastern Europe on average the students of female teachers tend to achieve higher test scores than of males. At the same time, the female teacher effect is more prevalent for girls. In Western Europe having a teacher of the same gender seems to benefit students.

In our interpretation, this pattern suggests that female teachers are more effective in general, while the teacher-student gender match might have a positive effect, as well. These effects cancel out when female teachers appear to have no effect on boys' test scores. However, in some countries there is no difference in the effect size by student gender, while in others the difference is statistically not significant, indicating that the student teacher gender match plays a lesser role. The gender match effect dominates the general female teacher effect only in England.

However, the teacher gender effect is far from universal; it is relatively strong in some countries while not present at all in many others. Cross-country differences in the gender gap in the relative teacher wages suggest selection into the teaching profession might be a major factor driving the female teacher effect. In countries where teacher wages relative to average graduate wages are higher for women than for men, teacher gender is more likely to affect student achievement.

Finally, there is a notable heterogeneity in the female teacher effect across various groups of students. Having a female teacher is associated with higher gains for low achievers, and in Western Europe for students with an immigrant background.

REFERENCES

Ammermüller, Andreas – Dolton, Peter (2006): Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA, *ZEW-Centre for European Economic Research Discussion Paper,* 06-060

Antecol, Heather - Eren Ozkan – Ozbeklik, Serkan (2015): The Effect of Teacher Gender on Student Achievement in Primary School, *Journal of Labor Economics*, 33, 1, pp. 63-89

Autor, David H. - Figlio, David N. – Karbownik, Krzysztof - Roth, Jeffrey - Wasserman, Melanie (2016): School Quality and the Gender Gap in Educational Achievement, *NBER Working Paper,* No. 21908

Bettinger, Eric P. - Long, Bridget T. (2005): Do faculty serve as role models? The impact of instructor gender on female students, *The American Economic Review,* 95, 2, pp. 152-157

Bland, J. Martin - Altman Douglas G. (1995): Multiple significance tests: The Bonferroni method, *British Medical Journal,* 310, 6973, p. 170

Carrington, Bruce - Tymms, Peter – Merrell, Christine (2008): Role models, school improvement and the 'gender gap'—do men bring out the best in boys and women the best in girls?, *British Educational Research Journal,* 34, 3, pp. 315-327

Chudgar, Amita - Sankar, Vyjayanthi (2008). The relationship between teacher gender and student achievement: evidence from five Indian states. *A Journal of Comparative and International Education*, 38(5), pp. 627-642

Clotfelter, Charles T. – Ladd, Helen F – Vigdor, Jacob L. (2006): Teacher-Student Matching and the Assessment of Teacher Effectiveness, *Journal of Human Resources*, 41, 4

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

Dee, Thomas S. (2005): A teacher like me: Does race, ethnicity, or gender matter?, *American Economic Review*, 95, 2, pp. 158-165

Dee, Thomas S. (2007): Teachers and the gender gaps in student achievement, *Journal of Human Resources,* 42, 3, pp. 528-554

Dolton, Peter - Marcenaro-Gutierrez, Oscar (2011): If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy*, 26 ,65, pp. 5-55

Dolton, Peter (2006): Teacher Supply, in: Hanushek Eric A. – Welch, Finis (eds.): *Handbook of the Economics of Education*, 2, pp. 1079–1161

Driessen, Geert (2007): The feminization of primary education: effects of teachers' sex on pupil achievement, attitudes and behaviour, *Review of Education*, 53, pp. 183–203

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

Eide, Eric - Goldhaber, Dan – Brewer, Dominic (2004): The teacher labour market and teacher quality, *Oxford Review of Economic Policy*, 20, 2, pp. 230-244

Escardíbul, Josep-Oriol – Mora, Toni (2013): Teacher gender and student performance in mathematics: Evidence from Catalonia, *IEB Working Paper*, 2013/7

Gershenson, Seth - Holt, Stephen B. - Papageorge, Nicholas W. (2015): Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations, *Institute for the Study of Labor*, Discussion paper, No. 9202

Hanushek, Eric A. – Rivkin, Steven G. (2006): Teacher quality, in: Hanushek Eric A. – Welch, Finis (eds.): *Handbook of the Economics of Education*, 2, pp. 1-25

Holmlund, Helena – Sund, Krister (2008): Is the gender gap in school performance affected by the sex of the teacher?, *Labour Economics,* 15, 1, pp. 37-53

Kane, Thomas. J.- Taylor, Eric S. - Tyler, John H. - Wooten, Amy L. (2011): Identifying effective classroom practices using student achievement data, *Journal of human Resources,* 46, 3, pp. 587-613

Krieg, John (2005). Student gender and teacher gender: what is the impact on high stakes test scores? Current Issues in Education 89

Lavy, Victor (2004): Do gender stereotypes reduce girls' human capital outcomes? Evidence from a natural experiment, No. w10678, *National Bureau of Economic Research*

Marks, Gary N. (2008): Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries, *Oxford Review of Education*, 34, 1, pp. 89–109

Muralidharan, Karthik - Sheth Ketki (2013): Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers, *NBER Working Paper,* No. 19341

Neugebauer, Martin - Helbig, Marcel – Landmann, Andreas (2011): Unmasking the Myth of the Same-Sex Teacher Advantage, *European sociological review*, 27, 5, pp. 669-689

Neugebauer, Martin – Helbig, Marcel – Landmann, Andreas (2011): Unmasking the myth of the same-sex teacher advantage, *European Sociological Review,* 27.5, pp. 669-689

Nixon, Lucia – Robinson, Michael (1999): The educational attainment of young women: Role model effects of female high school faculty, *Demography*, 36, 2, pp. 185-194

OECD (2015): Education at a glance 2015, *OECD*, Paris

OECD (2015): *The ABC of Gender Equality in Education: Aptitude, Behavior, Confidence*, OECD, PISA, OECD Publishing, ISBN 978-92-64-22994-5

Pekkarinen, Tuomas (2012): Gender differences in education, *Nordic Economic Policy Review*, 1

Rivkin, Steven G. – Hanushek, Eric A. – Kain, John F. (2005): Teachers, schools, and academic achievement, *Econometrica,* 73, 2, pp. 417-458

Terrier, Camille (2015): Giving a little help to girls? Evidence on grade discrimination and its effect on students' achievement, *Center for Economic Performance*, Discussion Paper, No. 1341

Varga, Júlia (2017): The Effects of Policies Concerning Teachers' Wages on Students' Performance, *BWP 2017/1*

Winters, Marcus A. - Haight, Robert C. - Swaim, Thomas T. - Pickering, Katarzyna 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, pp. 69-75

APPENDIX

Table A1

Sample size of students, teachers and classes by country

country	survey year	Student	student-subject	teacher	class	student, with teacher gender variation	share of students with teacher gender var.
west							
Bfl	1	4911	16528	580	269	2466	0,50
bsq	2	4696	9326	524	266	1463	0,31
eng	3	10102	18666	1852	524	4619	0,46
Fin	1	4266	19658	730	258	2867	0,67
Nld	1	3038	13156	436	129	1914	0,63
nor	3	12353	24342	806	600	2318	0,19
sco	2	7297	12358	1466	393	2714	0,37
swe	3	14328	37682	1622	821	5516	0,38
east							
bgr	2	7985	30384	1321	433	3126	0,39
bih	1	4220	20603	734	181	3592	0,85
cze	1	4817	23346	685	211	3361	0,70
est	1	3980	17392	663	153	2298	0,58
hun	3	12467	61022	2098	645	8584	0,69
ltu	3	13685	65846	2419	773	6672	0,49
lva	1	3592	13550	513	177	1514	0,42
mkd	2	7910	36513	1437	351	5346	0,68
rou	3	13825	68280	2578	692	10429	0,75
srb	2	8309	40349	1625	401	6305	0,76
svk	1	4215	17714	763	179	2447	0,58
svn	3	9031	39060	1998	516	5133	0,57

Table A2

Mean standardized math and science scores by student gender and country

country	math				science			
	boys		girls		boys		girls	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
west								
bfl	0.084	0.067	-0.060	0.068	0.282	0.054	-0.089	0.052
bsq	0.007	0.042	0.016	0.033	0.122	0.037	-0.102	0.032
eng	0.020	0.052	0.029	0.048	0.077	0.052	0.014	0.045
fin	-0.031	0.039	0.038	0.037	0.002	0.040	0.054	0.035
nld	0.049	0.088	-0.042	0.086	0.114	0.082	-0.135	0.080
nor	-0.021	0.023	0.028	0.021	0.014	0.023	-0.016	0.020
sco	0.014	0.052	0.038	0.049	0.119	0.051	-0.005	0.048
swe	-0.016	0.023	0.025	0.024	0.013	0.022	0.004	0.023
east								
bgr	-0.044	0.042	0.037	0.043	0.010	0.040	-0.017	0.042
bih	-0.011	0.044	0.003	0.045	0.013	0.043	-0.017	0.043
cze	-0.020	0.050	0.018	0.054	0.064	0.041	-0.072	0.048
est	-0.018	0.051	0.024	0.053	-0.023	0.044	0.031	0.046
hun	0.023	0.029	-0.032	0.030	0.118	0.026	-0.128	0.027
ltu	-0.048	0.026	0.042	0.024	-0.000	0.023	-0.012	0.022
lva	-0.053	0.051	0.033	0.046	0.022	0.046	-0.051	0.041
mkd	-0.062	0.044	0.037	0.043	-0.084	0.044	0.047	0.042
rou	-0.056	0.030	0.059	0.030	0.001	0.029	0.001	0.030
srb	-0.039	0.029	0.038	0.029	0.015	0.029	-0.012	0.026
svk	0.001	0.052	-0.001	0.048	0.115	0.046	-0.130	0.051
svn	-0.022	0.039	0.034	0.030	0.003	0.035	0.013	0.033

Notes: Science scores are averages of the scores for all science subjects. Robust standard errors are clustered at the class level.

Table A3

Mean standardized math and science scores by teacher gender and country

country	math				science			
	male		female		male		female	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
west								
bfl	-0.017	0.117	0.014	0.064	0.245	0.058	0.085	0.047
bsq	-0.003	0.063	0.018	0.036	0.062	0.060	-0.018	0.032
eng	-0.008	0.069	0.067	0.059	0.143	0.055	0.037	0.051
fin	0.016	0.051	-0.019	0.046	0.052	0.039	0.047	0.035
nld	0.028	0.099	-0.059	0.151	-0.025	0.077	0.002	0.100
nor	-0.018	0.027	0.028	0.027	-0.009	0.026	0.011	0.025
sco	-0.003	0.061	0.039	0.069	0.097	0.055	0.054	0.054
swe	0.021	0.029	-0.013	0.029	-0.002	0.025	0.014	0.025
east								
bgr	-0.057	0.095	0.004	0.042	0.011	0.049	-0.004	0.037
bih	-0.042	0.066	0.024	0.052	-0.039	0.040	0.005	0.041
cze	0.084	0.100	-0.025	0.057	0.032	0.054	-0.017	0.041
est	-0.143	0.116	0.021	0.051	0.050	0.050	0.006	0.041
hun	-0.157	0.060	0.028	0.030	-0.027	0.032	-0.003	0.025
ltu	-0.091	0.104	0.005	0.023	-0.028	0.030	-0.004	0.020
lva	-0.081	0.124	-0.002	0.048	-0.043	0.055	-0.004	0.040
mkd	-0.199	0.072	0.091	0.048	-0.177	0.045	0.048	0.041
rou	-0.011	0.046	0.011	0.034	0.000	0.032	0.008	0.027
srb	-0.099	0.042	0.068	0.033	-0.077	0.030	0.009	0.025
svk	-0.067	0.106	0.018	0.050	-0.035	0.049	0.003	0.044
svn	-0.150	0.075	0.038	0.036	-0.017	0.046	0.008	0.030

Notes: Science scores are averages of the scores for all science subjects. Robust standard errors are clustered at the class level.

Table A4

The estimated female teacher effect on test scores for low achievers and high achievers, Western Europe

	low achievers			high achievers		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bfl						
	0.0599***	0.0257	0.0841***	0.0288*	0.0271	0.0274
	(0.0226)	(0.0332)	(0.0257)	(0.0172)	(0.0207)	(0.0207)
bsq						
	0.0676**	0.0498	0.0887*	0.0371	0.0390	0.0368
	(0.0324)	(0.0427)	(0.0462)	(0.0350)	(0.0423)	(0.0437)
eng						
	0.00143	-0.0168	0.0160	-0.0147	-0.0317**	0.00251
	(0.0132)	(0.0181)	(0.0150)	(0.0108)	(0.0132)	(0.0150)
fin						
	0.00507	-0.0133	0.0273	0.00698	0.0121	0.00410
	(0.0169)	(0.0239)	(0.0179)	(0.0122)	(0.0147)	(0.0162)
nld						
	0.00619	-0.0200	0.0346	0.0177	0.0217	0.0128
	(0.0265)	(0.0295)	(0.0296)	(0.0249)	(0.0282)	(0.0303)
nor						
	0.0556**	0.0501*	0.0656**	0.0129	0.0353	-0.00913
	(0.0225)	(0.0295)	(0.0289)	(0.0217)	(0.0329)	(0.0268)
sco						
	-0.0138	-0.00790	-0.0190	0.0143	0.0140	0.0111
	(0.0195)	(0.0247)	(0.0253)	(0.0194)	(0.0277)	(0.0220)
swe						
	-0.00629	-0.0141	0.00364	0.0182	0.0130	0.0245
	(0.0142)	(0.0167)	(0.0182)	(0.0126)	(0.0172)	(0.0157)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Low/high achievers are students with average test scores over all subjects below/above the median student of their country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A5

The estimated female teacher effect on test scores for low achievers and high achievers, Eastern Europe

	low achievers			high achievers		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bgr						
	0.000205	0.00685	-0.00844	0.0540**	0.0781***	0.0268
	(0.0211)	(0.0236)	(0.0287)	(0.0220)	(0.0251)	(0.0274)
bih						
	0.00704	0.00487	0.0112	-0.0145	-0.0402*	0.0133
	(0.0188)	(0.0219)	(0.0233)	(0.0153)	(0.0209)	(0.0203)
cze						
	-0.0153	-0.0233	-0.00765	0.00511	0.0138	-0.00347
	(0.0144)	(0.0170)	(0.0207)	(0.0177)	(0.0198)	(0.0236)
est						
	0.0451*	0.0391	0.0490	0.0152	0.0181	0.0122
	(0.0231)	(0.0243)	(0.0360)	(0.0231)	(0.0285)	(0.0289)
hun						
	0.0139	0.00613	0.0205**	0.00643	-0.000934	0.0153*
	(0.00894)	(0.0123)	(0.0101)	(0.00745)	(0.0107)	(0.00896)
ltu						
	0.0153	0.00798	0.0236	0.00589	0.0113	0.00147
	(0.0128)	(0.0158)	(0.0175)	(0.0102)	(0.0128)	(0.0138)
lva						
	0.00570	-0.00108	0.00978	-0.0278	-0.0190	-0.0376
	(0.0250)	(0.0361)	(0.0287)	(0.0266)	(0.0330)	(0.0316)
mkd						
	0.0140	0.00588	0.0241	0.0271**	0.0198	0.0332**
	(0.0120)	(0.0159)	(0.0152)	(0.0135)	(0.0171)	(0.0147)
rou						
	0.00535	0.00130	0.00938	0.0107	-0.00354	0.0262*
	(0.00983)	(0.0116)	(0.0135)	(0.0109)	(0.0136)	(0.0138)
srb						
	0.0185*	0.0112	0.0255*	0.0249**	0.0203	0.0298**
	(0.0103)	(0.0130)	(0.0144)	(0.0111)	(0.0140)	(0.0144)
svk						
	4.86e-05	-0.0426*	0.0370	0.0194	0.0113	0.0317
	(0.0210)	(0.0247)	(0.0268)	(0.0225)	(0.0265)	(0.0301)
svn						
	0.0164	0.0308	0.000325	-0.0184	-0.00856	-0.0271
	(0.0166)	(0.0230)	(0.0197)	(0.0144)	(0.0181)	(0.0185)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Low/high achievers are students with average test scores over all subjects below/above the median student of their country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A6

The estimated female teacher effect on test scores for students with few and many books at home, Western Europe

	few books			many books		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bfl						
	0.0394**	0.0238	0.0551***	0.0442**	0.0261	0.0580*
	(0.0156)	(0.0193)	(0.0206)	(0.0224)	(0.0272)	(0.0307)
bsq						
	0.0656*	0.0245	0.110**	0.0356	0.0619	0.00705
	(0.0334)	(0.0374)	(0.0504)	(0.0294)	(0.0409)	(0.0356)
eng						
	-0.00876	-0.0235*	0.00538	0.00240	-0.0223	0.0279*
	(0.0115)	(0.0141)	(0.0140)	(0.0107)	(0.0154)	(0.0150)
fin						
	-0.00302	-0.000181	-0.00192	0.00845	-0.00984	0.0206
	(0.0133)	(0.0175)	(0.0184)	(0.0133)	(0.0178)	(0.0171)
nld						
	0.0220	0.00156	0.0464	-0.000839	-0.00246	0.00508
	(0.0260)	(0.0289)	(0.0302)	(0.0269)	(0.0300)	(0.0313)
nor						
	0.0586***	0.0452	0.0733***	0.0119	0.0560	-0.0167
	(0.0208)	(0.0282)	(0.0283)	(0.0258)	(0.0399)	(0.0313)
sco						
	-0.0154	-0.00841	-0.0235	0.0313	0.0164	0.0431*
	(0.0163)	(0.0228)	(0.0211)	(0.0235)	(0.0387)	(0.0257)
swe						
	-0.00240	-0.00436	-0.000256	0.0195	0.00835	0.0291*
	(0.0133)	(0.0153)	(0.0181)	(0.0128)	(0.0193)	(0.0166)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Few/many books refers to the number of books at home below or equal to / above the median value of the country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A7

The estimated female teacher effect on test scores for students with few and many books at home, Eastern Europe

	few books			many books		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bgr						
	0.0129	0.0258	-0.00246	0.0336*	0.0474**	0.0197
	(0.0234)	(0.0253)	(0.0299)	(0.0188)	(0.0215)	(0.0338)
bih						
	-0.00605	-0.0144	0.00400	0.00555	-0.0308	0.0347
	(0.0167)	(0.0203)	(0.0202)	(0.0180)	(0.0220)	(0.0258)
cze						
	-0.000424	-0.00394	0.00457	-0.00624	0.00705	-0.0164
	(0.0119)	(0.0136)	(0.0194)	(0.0185)	(0.0237)	(0.0228)
est						
	0.0183	0.00609	0.0292	0.0449**	0.0515*	0.0360
	(0.0216)	(0.0240)	(0.0310)	(0.0218)	(0.0273)	(0.0299)
hun						
	0.00916	0.00274	0.0158*	0.00821	-0.000381	0.0166
	(0.00800)	(0.0106)	(0.00940)	(0.00841)	(0.0126)	(0.0105)
ltu						
	0.0142	0.0143	0.0145	-0.00122	-0.00907	0.00663
	(0.0101)	(0.0124)	(0.0144)	(0.0144)	(0.0190)	(0.0179)
lva						
	-0.0245	-0.0153	-0.0339	0.0107	0.00684	0.0192
	(0.0223)	(0.0291)	(0.0260)	(0.0336)	(0.0446)	(0.0395)
mkd						
	0.0233*	0.0177	0.0311*	0.0223*	0.0164	0.0281**
	(0.0125)	(0.0150)	(0.0168)	(0.0128)	(0.0170)	(0.0140)
rou						
	0.00567	-0.000896	0.0128	0.0165	-0.00630	0.0366**
	(0.00912)	(0.0108)	(0.0121)	(0.0118)	(0.0156)	(0.0153)
srb						
	0.0225**	0.0126	0.0327**	0.0165	0.0147	0.0175

	(0.00941)	(0.0124)	(0.0130)	(0.0113)	(0.0142)	(0.0144)
svk						
	-0.00150	-0.0324	0.0330	0.0372	0.0231	0.0526
	(0.0206)	(0.0222)	(0.0291)	(0.0240)	(0.0332)	(0.0347)
svn						
	0.00982	0.0252	-0.00728	-0.0444**	-0.0559*	-0.0348
	(0.0133)	(0.0154)	(0.0183)	(0.0197)	(0.0286)	(0.0277)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Few/many books refers to the number of books at home below or equal to / above the median value of the country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A8

The estimated female teacher effect on test scores for students with immigrant and non-immigrant background, Western Europe

	immigrant			non-immigrant		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bfl						
	0.0367	-0.00323	0.0820	0.0433***	0.0272	0.0564***
	(0.0347)	(0.0436)	(0.0503)	(0.0153)	(0.0185)	(0.0185)
eng						
	0.000760	-0.0262	0.0244	-0.0120	-0.0272**	0.00213
	(0.0178)	(0.0212)	(0.0221)	(0.00939)	(0.0125)	(0.0122)
nld						
	0.0447	0.0287	0.0652	0.00285	-0.0104	0.0184
	(0.0389)	(0.0502)	(0.0458)	(0.0198)	(0.0221)	(0.0227)
nor						
	0.0256	-0.0630	0.109***	0.0304	0.0516**	0.00890
	(0.0312)	(0.0475)	(0.0384)	(0.0193)	(0.0260)	(0.0257)
swe						
	0.0128	-0.00545	0.0334	0.00461	0.00106	0.00828
	(0.0154)	(0.0209)	(0.0220)	(0.0110)	(0.0141)	(0.0129)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Immigrant background refers to either student or mother or father born in another country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A9

The estimated female teacher effect on test scores for students with immigrant and non immigrant background, Eastern Europe

	immigrant			non-immigrant		
	FE full	FE boys	FE girls	FE full	FE boys	FE girls
	(1)	(2)	(3)	(4)	(5)	(6)
bih						
	-0.0165	-0.0224	-0.00918	-0.00169	-0.0194	0.0165
	(0.0217)	(0.0274)	(0.0305)	(0.0142)	(0.0176)	(0.0183)
est						
	0.0427	0.0332	0.0485	0.0234	0.0216	0.0237
	(0.0304)	(0.0308)	(0.0510)	(0.0200)	(0.0239)	(0.0247)
lva						
	-0.00942	-0.000359	-0.0176	-0.0106	-0.00598	-0.0156
	(0.0320)	(0.0442)	(0.0413)	(0.0205)	(0.0322)	(0.0227)
srb						
	0.0153	0.00214	0.0311	0.0227**	0.0180	0.0270**
	(0.0155)	(0.0187)	(0.0221)	(0.00887)	(0.0109)	(0.0119)
svn						
	-0.0189	0.0208	-0.0579**	0.000622	0.00274	-0.00267
	(0.0200)	(0.0294)	(0.0252)	(0.0147)	(0.0194)	(0.0157)

Notes: Dependent variable is standardized test score. Each cell presents a coefficient for a separate regression estimate of having a female teacher. Immigrant background refers to either student or mother or father born in another country. Robust standard errors clustered at the class level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1