Better test scores with a same-gender teacher?*

Johan Coenen
Chris Van Klaveren†

Abstract

The literature on gender specific educational achievement indicates that primary school teaching has become more and more a female profession and that the lack of male role models in primary education may negatively influence the school achievement of boys. This study examines if children's math test scores are higher if they are taught by a teacher of their own gender. For this purpose we use unique Dutch data on 2586 primary school children and identify the same-gender effect by estimating an innovative within class between child gender estimation model. The empirical findings indicate that children's math performance is not influenced by the gender of their teacher.

JEL-codes: I20 · I21 · I29

Keywords: Teacher · Student achievement · Gender · Mathematics

*We would like to thank Torberg Falch, Wim Groot, Henriette Maassen van den Brink, the participants from the ESPE 2012 conference, IWAEE 2012 and the SES conference 2012 for their valuable comments.

†Johan Coenen (Corresponding author), TIER, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. E-mail: johan.coenen@maastrichtuniversity.nl. Chris Van Klaveren, TIER, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. E-mail: cp.vanklaveren@maastrichtuniversity.nl.

1 Introduction

The empirical literature on gender specific educational achievement in primary schools shows that boys outperform girls in mathematics and science subjects, but that the test scores differences between boys and girls decrease (or reverse) over time (Hedges and Nowell 1995; Willingham and Cole 1997; Neugebauer, Helbig and Landmann 2010). At the same time, this literature shows that girls tend to outperform boys in language related subjects. These achievement differences have long been considered as given, but the diminishing (and sometimes reversed) achievement gaps, together with the observed higher share of girls in higher secondary education levels and of women among first-year students in Higher Education have changed the traditional expectations about gender differences in education (see, among others, Pollack 2006; Neugebauer, Helbig and Landmann 2010; OECD 2010).

The observed gender specific achievement patterns observed in primary education caused a debate in which two arguments are central. The first argument is that gender specific achievement differences are worrisom because of the possible long-term effects that these differences may have on future labor market and educational outcomes (Merrell and Tymms 2011). The second argument is that primary school teaching has become more and more a female profession and this might have contributed to the observed gender specific achievement patterns. In some studies it is argued that the feminization of the primary school teaching profession has led to a lack of male role models, which negatively affects the achievement of boys (Driessen 2007; Holmund and Sund 2008). Other studies show that teachers assess and/or grade children of the same gender differently than children of the opposite gender. Given the large share of female teachers in primary education, this may negatively affect the achievement of boys (Ouazad 2008; Ammermüller and Dolton 2006; Mechtenberg 2009; Falch and Naper 2011).

This study is related to the feminization of the primary school teaching profession and examines, in particular, if children's math test scores are higher if they have been taught by a teacher of their own gender. For this purpose unique registration data is used that contains background and test score information of 2586 Dutch primary school children from the grades 3 to 5. In addition to this, data from our questionnaires, furthermore contain information on the teachers and parents of these children and contain information on how teachers and children are assigned to classes.

We proceed as follows. Section 2 describes the literature on gender specific educational achievement differences. Section 3 discusses the identification and estimation strategy. Section 4 describes the data and show the descriptive statistics and Section 5 presents the empirical results. Section 6 investigates how robust the results are, and examines among others if (fe)male teachers select themselves in better classes, and to what extent the teacher gender and average class math score of the previous school year affect the results. Finally, Section 7 concludes.

2 Literature

Traditionally, there have been gender differences in school performance. While boys in general performed better in mathematics and science, girls proved to be the better in language and reading. A wide range of studies is performed to explain gender differences in educational attainment in general, with some studies focusing on the reversed gender gap: girls outperforming boys. Buchmann, DiPrete and McDaniel (2008) have reviewed the literature on gender differences in educational attainment, for primary education and secondary education on the one hand, but also on gender differences in early adulthood, concerning school completion and enrollment in post-secondary education. In their literature review, they have found gender inequalities in education throughout the entire educational career. Already in kindergarten, there are differences in level between boys and girls. For instance, Graue and DiPerna (2000) have found that in the US, the occurrence of delayed entry into kindergarten is higher among boys than among girls. Malone et al. (2006) have found as well that more boys start kindergarten at an older age, and furthermore have found boys to be 66% of all children who repeated kindergarten. Alexander et al. (2003) and Entwisle et al. (2007) have concluded that boys are more likely to be retained one or more grades during elementary education. These facts have got consequences for performance comparisons between boys and girls. When comparing at the grade level, boys will be on average a bit older. When comparing at the age level, girls on average will have attained a bit higher grade level. Besides this, it is often argued that girls tend to mature faster than boys, which can lead to an additional developmental advantage of girls over boys (Buchmann, DiPrete and McDaniel 2008). Despite the many studies that have examined the gender education gap, disagreement on some important issues remains. There is, for instance, disagreement on at what age performance differences in mathematics between boys and girls emerge (Leahey and Guo 2001), whether the gender gap in test scores narrows over time (see, among others, Feingold 1988; Hyde et al. 1990; Hedges and Nowell 1995), and whether the test score variation for boys is larger (Willingham and Cole 1997).

Older studies by Maccoby and Jacklin (1974), and by Willingham and Cole (1997) point however at similar performance of boys and girls on mathematics and reading in the earliest grades, with a growing advantage in math for boys, and a growing advantage in reading for girls when pupils move to higher grades. There have been some studies which suggest that the gender education gap in test scores is wider among children from parents with a low SES (Hinshaw 1992). Entwistle et al. (2007) have indeed found that a reading gap in favor of girls only emerged among children from parents with low income.

The discussion about the feminization of education is already old. Ever since teaching became a predominantly female profession, concerns were raised about boys lacking male role models in schools. Apparently, the reason for these concerns were a fear that boys would not develop enough masculinity (Sexton 1969). Back then, but also nowadays, there was and is a striking focus on boys. A fact, which is often neglected, is that girls outperforming boys, does not per se mean that the educational attainment of boys has decreased. In many cases, the educational attainment of girls has risen. In that regard, it might not be the failure of boys, but the success of girls, which explains the reversed gender gap in education. Female teachers outnumbered male teachers in primary education already decades ago, but the percentage of male teachers has declined even further in recent years (Corcoran, Evans and Schwab 2004).

So why is the feminization of the teaching profession regarded as a bad thing? The first argument is already mentioned in the previous paragraph: there would be not enough male role models in school for boys to identify with (Driessen 2007; Holmund and Sund 2008).

There are some studies in the literature on possible gender interaction effects, that is, boys profiting from having a male teacher, and girls profiting from having a female teacher. These studies vary in multiple ways. Some studies use subjective teacher assessments as variable of interest. These could, however, measure teacher favoritism for students from the same gender. Other studies look at effects on blind (national) test scores. A third kind of studies uses data from international comparable tests, like PIRLS for literacy and reading and/or TIMMS for mathematics and science (See for instance Ammermüller and Dolton 2006). Some studies combine blind test scores with teacher assessment (Dee 2007; Neugebauer, Helbig and Landmann 2010). These different approaches to comparing learning performance of boys and girls are not trivial. Duckworth and Seligman (2006) conclude that boys tend to get higher scores in standardized tests, while girls receive higher grades in class. The latter is not a new phenomenon. Back in the fifties and sixties, girls already received higher grades than boys (Alexander and Eckland 1974; Alexander and McDill 1976; Mickelson 1989). Perkins et al. (2004) conclude that nowadays girls get higher grades than boys in every major subject, on all levels in education, from kindergarten to college.

Results from earlier studies show mixed results. Dee (2007) has examined whether assignment to a same-gender teacher influenced student achievement, teacher perceptions of student performance, and student engagement. He concludes that within-student comparisons indicate that assignment to a same-gender teacher significantly improves the student achievement of both boys and girls, as well as teacher perceptions of student performance, and student engagement with the teacher's subject. He has analyzed this by differencing two separate equations, in which the educational outcome of a student in a subject is a function of observed student traits, and whether the teacher of the class is female. Neugebauer, Helbig and Landmann (2010) on the other hand, have not found an effect of the teacher's gender on learning outcomes at all. They have found that boys do not benefit from male teachers and girls do not (significantly) benefit from female teachers, neither on blind test scores, nor on grades given by the teachers. Ammermüller and Dolton (2006) have found mixed results in their study. They have used international test score data for reading, mathematics, and science, from respectively PIRLS, and TIMMS, for students of grade 4, and grade 8, from the US, and the UK. They have found positive male interaction effects in mathematics scores in the US, and in science scores in England, at grade 8. Furthermore, using individual fixed effects, they have found positive joint pupil-teacher gender interaction effects in mathematics, for both boys and girls of grade 8, but only in England, not in the US, and only in 2003, not in 1995 or 1999. They have not found large performance differences between boys and girls in TIMMS, as opposed to the national SAT, or state wide tests in the US, and the GCSE test in the UK. Therefore, they suggest that these country specific exam systems favor girls and the way they learn and study. On the other hand, the characteristics of the international and the national test are quite different, and also taken at different ages.

Holmlund and Sund (2008) have studied whether same gender teachers have improved the grades of academic secondary education students in Sweden. After controlling for a different teacher gender composition in different subjects, they have found a larger gender performance gap in subjects with a higher proportion of female teachers. They argue however, that this larger gender performance gap is due to teacher selection into different subjects, and non-random assignment of students to teachers. Therefore, they have investigated the within-student and subject effect, using teacher turnover or student mobility. In this specification with student fixed effects, they have found practically no significant effects from having a same gender teacher. Sokal et al. (2007) have also found no effects of same gender teachers on the reading performance of boys. In Canada, they have conducted an intervention of 10 weeks extra reading instruction for boys, who were performing badly in reading in schools. Although they have not found any effects of teacher gender on reading performance, they have showed how the intervention has changed boys' perception of reading in a good way. Earlier, we have mentioned already how part of a potential same gender effect could be due to teacher discrimination; in this case, teachers favoring students of the same gender. Ouazad (2008), in particular, has investigated whether teachers discriminate in their assessments of their students. Using longitudinal data from elementary education in the US, which features both test scores and teacher assessments, he has analyzed whether discrimination based on gender, but also on race and ethnicity, takes place. He has found evidence of discrimination by teachers, but not for gender-based discrimination. Teachers gave higher assessments to students of the same race. A result which seems to be primarily driven by white teachers giving a lower assessment to non-Hispanic black children and Hispanic children.

3 Identification and Estimation Strategy

This section outlines the strategy to identify if there is a same-gender teacher effect on math performance. We can estimate the association between teacher characteristics and math performance by estimating the following education production function:

$$y_{ics}^{end} = \alpha_0 + \alpha_1 X_i + \alpha_2 T_{ic} + \alpha_3 C_i + \alpha_4 S_{is} + \varepsilon_{ics}, \tag{1}$$

where y_{ics}^{end} represents the math performance at the end of the school year for pupil i in class c and school s that depends on pupil (X), teacher (T), class (C) and school (S) characteristics. As is usual, the error term, ϵ_{ics}, is assumed to be normally distributed with mean zero and variance σ_{ϵ}^2 and all explanatory variables are assumed independent of the error term. If we would estimate equation 1 and include a dummy variable, I, that indicates if children are being taught by same-gender or opposite-gender teacher, we would not take into account selection effects that might occur. Therefore the estimated coefficient for I cannot be interpreted as an *effect*, because pupil, teacher and school characteristics, that are partly unobserved, could be correlated with both y and I, and could occur systematically in the error term, which leads to a bias in all the parameter estimates (see Van Klaveren (2011) who describes the selection problem using a similar context and by providing a similar identification method). The estimates are, for example, biased if high ability children go to better schools and if, at the same time, more female teachers work on these better schools.

To account for selection and omitted variable bias and to quantify the same-gender teacher effect we adopt a two-step approach. In the first step we take into account that children bring different levels of achievement, or knowledge, to the classroom at the beginning of the school year and that this influences that estimated teacher effects. To control for these achievement-level differences at the beginning of the school year, we adopt a within-class matching approach where boys are paired with girls based on the math test scores achieved at the end of last school year. In the second step we identify the same-gender teacher effect for these paired child couples by estimating a within class between child gender estimation model.

Pairing boys and girls within classes

For each class we pair boys and girls within classes conditional on the observed math test scores at the end of last school year (hereafter referred to as math pretest scores). Let B and G denote the set of boys and girls in the class and let $y_{pre,B}$ and $y_{pre,G}$ be the normalized math pretest scores for respectively boys and girls. We can determine which boys and girls are most similar in their math pretest scores by calculating the normalized Euclidean distance between each boy and each girl in the following way:

$$D_{b,g} = \sqrt{(y_{pre,b} - y_{pre,g})^2 + (\bar{y}_{G,pre} - \bar{y}_{G,pre})^2}. \tag{2}$$

The smaller the distance that we observe between boy b and girl g the more similar they are in terms of their math pretest score. We note that the $D-$values are normalized, even though equation 2 represents the unnormalized Euclidean distances, because y_{pre} are normalized pretest scores.

If a class consists of more girls than boys we match girls to boys following an iterative two-step procedure. In the first step, we create for each boy an ordered list of girls conditionally on the calculated $D-$values. In the second step we select the boy-girl couple with the lowest $D-$value and remove the paired girl from the created lists of all other boys. We repeat these two steps until all boys are paired with a girl. A similar procedure is followed if a class consists of more boys than girls.

This iterative two step procedure results in within-class samples of boys and girls with more comparable math pretest scores and thereby we (partially) control for the fact that children may bring different levels of achievement to the classroom at the beginning of the school year. An additional advantage of this pairing mechanism is, first of all, that the endogenous math pretest score variable does not have to be included as explanatory variable in the empirical model described in 1. Another advantage is that the pairing mechanism is nonparametric such that no *a priori* assumption are made about how the achieved math test scores at the end of the school year depend on the math pretest scores (Yatchew 1998).

A disadvantage of the pairing mechanism is that the $D-$values increase with the number of iterations. We therefore perform the empirical analysis for all boy-girl pairs, but also perform the analysis when we consider only the pairs with, respectively, the lowest 25, 50 and 75 percent $D-$values, i.e. the best matches. Additionally, we perform a matching approach where we allow girls (boys) to be matched multiple times to boys (girls). This empirical analysis for the matched boys and girls with replacement serves as a robustness check and examines how the empirical results change if math pretest scores between boys and girls are more similar.

The within class between child gender estimation model

The within class between child gender estimation model considers the test-score differences of the paired boys and girls at the end of the school year, such that equation 1 can be rewritten as:

$$\begin{aligned}\Delta y_{end,k} = y_{end,G_k} - y_{end,B_k} = (\alpha_{0,G_k} - \alpha_{0,B_k}) + \Delta X^{'}_{G-B,k}(\alpha_{1,G_k} - \alpha_{1,B_k}) \\ + T^{'}_k(\alpha_{2,G_k} - \alpha_{2,B_k}) + C^{'}_k(\alpha_{3,G_k} - \alpha_{3,B_k}) + S^{'}_k(\alpha_{4,G_k} - \alpha_{4,B_k}) + \nu_k,\end{aligned} \tag{3}$$

$\Delta y_{end,k}$ represent the test score differences for the k couples at the end of the school year. Subscript $G - B$ indicates that we always subtract the boys characteristic from the girl characteristic, and subscript k refers to a particular boy-girl pairing. There is an abundant literature on the existing performance gap between boys and girls that suggest that boys outperform girls on math, and that girls outperform girls on reading (see for instance Marks 2008; Buchmann, DiPrete, and McDaniel 2008). Therefore y_{end,G_k} and y_{end,B_k}represent the standardized values to control for constant outperforming differences between girls and boys. The math performance of girls and boys is probably influenced in a similar way by class and school characteristics, which implies that $\alpha_{3,G_k} - \alpha_{3,B_k} = 0$ and $\alpha_{4,G_k} - \alpha_{4,B_k} = 0$. A similar argument holds for teacher characteristics, such as teacher education and experience.

Teacher gender, however, influences the performance of boys differently than the performance of girls. In classes with a female teachers it is automatically the case that girls have a same-gender teacher, and that boys have an opposite-gender teacher, while the opposite is true for classes with male teachers. Because $\Delta y_{G-B,k}$ subtracts test scores of boys from the test scores of girls, the same-gender teacher effect can be measured by including an indicator variable, $TFEM$, that equals 1 if the teacher is a woman, and 0 otherwise. Test score differences between paired girls and boys may also depend on differences in their background characteristics, $\Delta X_{G-B,k}$, and so these background characteristics are included in the model as well. It follows that equation 3 can be rewritten as[1]:

$$\Delta y_{G-B,k} = \gamma_0 + \Delta X'_{G-B,k}\gamma_1 + TFEM'_k\gamma_2 + \nu_k \tag{4}$$

Because the interpretation of the same-gender teacher effect, γ_2, is not straightforward, we explain in detail how the same-gender teacher effect can be identified based on the estimated value for γ_2. In the exposition below we assume that the girls and boys within each couple have identical math pretest scores. For the identification of the same-gender teacher effect, this assumption is crucial, and in Section 4 we show the extent to which this assumption is realistic.

Table 1: Achieved math test scores by child and teacher gender

	Child Gender		
$TFEM$	Girls	Boys	Δ
1	y_1^1	y_0^1	$\Delta_1 = y_1^1 - y_0^1$
0	y_1^0	y_0^0	$\Delta_2 = y_1^0 - y_0^0$

Table 1 shows the math test scores achieved at the end of the school year by teacher and child gender. The first column indicates if a child has a female teacher ($TFEM = 1$), or a male teacher ($TFEM = 0$). The second and third columns show the achieved test scores for girls and boys, y. The superscript indicates 1 if the child has a female teacher, and 0 otherwise. The subscript indicates 1 if the child is a girl, and 0 otherwise. The last column presents the performance differences between girls and boys separately for female teachers (Δ_1) and male teachers (Δ_2). Table 1 is used to explain how the same-gender teacher effect is identified under the assumptions that (1) matched boys and girls perform equally well at the beginning of the school year, and (2) that female and male teachers teach to comparable children in terms of pretest scores and background characteristics. We return to the credibility of the identifying model assumptions in Section 4.

Table 1 shows that girls with a female teacher perform better than boys if $y_1^1 > y_0^1$ which results in a positive Δ_1. Boys with a male teacher perform better than girls if $y_0^0 > y_1^0$, which results in a negative value Δ_2. The empirical model links Δ - differences to teacher gender, and generates positive γ_2 values if $\Delta_1 > \Delta_2$, and thus if girls structurally perform better than boys, if they are being taught by a female teacher. It is however possible that girls perform better than boys, if they are being taught by a female teacher ($y_1^1 > y_0^1$),

[1] As a robustness check we estimate both equations 3 and 4.

and that girls perform better than boys, if they are being taught by a male teacher ($y_0^0 < y_1^0$), or vice versa. In this particular situation, the empirical model will generate positive $\gamma_2$values as long if $\Delta_1 > \Delta_2$ (i.e. if the same-gender effect exceeds the opposite-gender effect). If there is an opposite-gender effect, or if the opposite-gender effect exceeds the same-gender effect, then Δ_1 will be smaller than Δ_2, which will translate in a $\gamma_2$estimate that is negative.

Finally, it may be that structural differences between Δ_1 and Δ_2 are not related to the gender of the teacher, which results in an estimated value of γ_2 that is not significantly different from zero. This non-significance indicates, (1) that there are no opposite-gender or same-gender teacher effects, or indicates (2), that there is an opposite-gender effect for male teachers, which is equal to same-gender effect for female teachers (or vice versa). There is a large literature that argues that the latter situation is highly unlikely (Ammermüller and Dalton 2006; Dee 2007; Neugebauer, Helbig and Landmann 2010), but we may not exclude this possibility a priori. As a robustness check we therefore characterize Δ_1 and Δ_2 based on the observed y_1^0, y_0^0, y_1^1, and y_0^1, and examine if there are structural differences between Δ_1 and Δ_2 that are related to teacher gender.

4 Data and descriptive statistics

This paper uses unique panel data on pupils of Dutch primary schools. These data come from a three year long (and still ongoing) field experiment, conducted to evaluate the effectiveness of various school time extension programs (see for example, Meyer and Van Klaveren 2011). We make use of different data sets, which were all collected primarily to evaluate an experiment in primary education: the extended school day. Schools, which offer the extended school day concept, provide their pupils with increased instructional time, in an attempt to decrease their learning deficiencies. For the evaluation, both schools who offer the extended school day, and schools who do not offer it, are in the evaluation study. The study consists of: (1) a questionnaire for the pupils, (2) a questionnaire for the pupil's parents, (3) a questionnaire for the teachers, and (4) a questionnaire for the school directors. The before mentioned research project does not focus on a link between having a same-gender teacher and learning achievement. The data, however, allow for a research extension in this direction. For the sake of this project, information is obtained on teacher characteristics, such as experience, education level, wage, gender, and ethnicity. Moreover, we have obtained detailed information on pupils and their parents, that originates partly from the school registration system, and partly from pupil and parent questionnaires. We focus on pupils of the grades 3, 4, and 5 (pupils aged approximately 7 to 11 years old)[2]. For this study, we merged the resulting data sets. We linked the pupil (and parents) data with the teacher data, by tracing each teacher who taught the particular classes of the pupils from the participating schools. Once we were able to link pupils to their respective teacher or teachers, we merged these linked pupils with the CITO-LOVS data set, which contains test scores for most pupils. Mathematics performance is measured as the test scores that children achieved on this LOVS test, a national and standardized math test that is taken bi-annually by their schools (i.e. in February and in June). We use the test scores from June 2010 as pretest score, and the test scores from June 2011 as post test score. Pupils for which we could not find the corresponding teacher, or pupils without available pretest

[2]The test is also taken by pupils in earlier grades, but we can only use pupils from the grades 5 to 7, because earlier grades are not in the research project. pupils from grade 8 do not take two tests, only the final CITO test, which partly decides to what level of secondary education they can go after finishing primary education.

Table 2: Descriptive statistics for male and female teachers

	Male teachers	Female teachers
Teacher		
Total work experience (in years)	19.6	14.0
School		
Total pupils in school (#)	181	160
School in poverty problem area (%)	43.7	44.8
Children from parents with very low or no education (%)	18.4	15.9
Children from parents with low education (%)	13.4	15.2
Boys in school (%)	51.0	50.1
Pupils from immigrant background at school (%)	52.9	50.0
Class		
Boys in class (%)	50.7	48.8
Class size (#)	21.1	20.7
Pupil		
Education level mother (of pupils in class) (%)		
Unknown	4.7	8.2
No, primary or lower vocational education	17.7	13.4
Higher secondary education or Intermediate vocational education	52.3	58.2
Higher vocational education or University	25.4	20.2

and posttest scores, are excluded from the analysis.

Our data is in no way a representative sample of Dutch primary education pupils. The schools in our sample, in general, have a very high proportion children from a low social background, and a very high proportion children with a non-Dutch background. This is related to the fact that our data is based on the evaluation of an experiment which has the purpose of decreasing educational deprivation. Obviously, such experiment will be most likely to be held on schools with many pupils from a low social background.

We arranged the data on a class level. Some teachers were excluded, because we had no information about their classes, and some classes were excluded, because we had no information about their teachers. We use 179 classes from 76 schools, which have a total of 174 teachers and 2586 pupils. We could not use many pupils, because they had missing information on their teacher gender, or missing test scores. Pupils need to have two test scores, otherwise we either cannot match pupils together based on pretest, or we cannot include them in the analysis, which has differences in posttest scores of matched girls and boys as dependent variable. Every pupil has separate values for the pupil, test scores, and parents data, but shared data for the teacher. Some teachers teach more than one class at the same time.

In Table 2 some descriptive statistics for a selection of the observables are shown. Male and female teachers are compared on some teacher, school, class, and pupil characteristics. These descriptive statistics are based on the sub-sample we use in the analysis.

Table 2 shows that there are some differences between male and female teachers concerning their observable school, class, pupil and own background characteristics. The male teachers in our data have on average 5.6 more years of working experience than the female teachers. Male teachers work on average at slightly larger schools, i.e. schools with more children. Male and female teachers work to a similar extent at schools in poverty problem areas[3], but male teachers work at schools with a higher level of pupils with very low, or uneducated parents[4], while female teachers work at schools with a slightly higher level of pupils with low educated parents[5]. The percentages of boys, both in school and in the class, is similar for male and female teachers, but male teachers work on average at schools with a slightly higher level of pupils from a non-Dutch background. Both male and female teachers teach classes with on average 21 pupils. In comparison to female teachers, male teachers have on average more pupils in their class with higher educated parents, but also more pupils with parents which have no education, primary education, or lower vocational education as highest completed education level.

Next, we will investigate the pretest scores for the matched girl-boy couples for classes with female teachers and classes with male teachers. As we emphasized in Section 3, our identification strategy relies on the assumption that the pretest scores of the boys and girls in the matched couples are equal. To test this assumption, we check separately for classes with female and male teachers, if the differences in pretest scores between boys and girls in couples are significantly different from each other. In Table 3 the pretest scores are shown for girls and boys, separately for female and male teachers. the mathematics pretest scores in Table 3 are the LOVS end of the school year test scores from June 2010. In all samples, the number of boys and girls is half of the total number of pupils.

After matching, our data contains 2586 pupils, or 1293 couples, the before mentioned boy-girl pairings within a class based on a similar pretest score. Ideally, we would have pretest scores which are equal within the matched couple: zero difference in test scores from June 2010 between the matched girl and boy. With all other (non-pupil) characteristics equal for the boy and the girl in a class, besides whether the teacher is a same-gender, or an opposite-gender teacher, differences in post test scores can then solely be ascribed to a same-gender, or opposite-gender teacher effect. In practice, we still find some differences between boys and girls in pretest scores. Furthermore, pupils from male teachers and female teachers on average do not have the same pretest score. We can conclude some points from Table 3: (1) On average, boys performed better in the mathematics pretest than girls. (2) On average, pupils from male teachers score higher mathematics test scores than pupils from female teachers. With the full sample, the difference between boys and girls with a male teacher is, however, a non-significant 1.35 points. For couples with a female teacher this difference is larger and significant: 2.96 points. (3) For the best 75% matches, the difference become smaller: now the children of the female teachers only have a 1.62 difference in pretest score. This difference is still significant on the 10% level though. (4) For the best 50%, the best 25% matches, and for the matches made with replacement, we no longer find significant differences between the test scores of the matched couples. Since we already do not find significant differences between the couples for the best 50%, we will not use the best 25% samples, because this leads to a smaller sample size. Therefore, in the rest of the paper, we will show the results for the analysis for the full sample, the best 75% matches, and the best 50% matches.

[3]Areas defined by Statistics Netherlands as postal codes regions with at least 8.6% of all households with a low income and at least 7.2% of all main providers from a non-Dutch ethnic background

[4]Maximum education level of 1 or both parents: primary education or special education. Schools get additional funds to reduce educational deprivation, based on the percentage of these pupils.

[5]Maximum education level of both parents, or 1 parents which is the guardian: lower vocational education or practice education. Schools get additional funds to reduce educational deprivation, based on the percentage of these pupils.

Table 3: Average mathematics pretest scores for boys and girls with male and female teachers, for 25% to 100% of the matches

	100%		75%		50%		25%		Replacement	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Female Teachers										
Girls	75.76	(18.09)	76.45	(17.96)	76.60	(18.00)	78.00	(17.93)	77.11	(17.76)
Boys	78.72	(17.69)	78.07	(17.85)	77.92	(17.99)	78.71	(17.31)	77.86	(17.92)
Difference Girls - Boys ($\Delta_1 = y_1^1 - y_0^1$)	-2.97***	(0.78)	-1.62*	(0.94)	-1.32	(1.16)	-0.71	(1.73)	-0.75	(0.78)
Number of teachers	141		141		141		128		141	
Number of pupils	2090		1456		962		414		2090	
Male Teachers										
Girls	77.67	(17.66)	77.99	(16.97)	79.11	(16.98)	78.90	(18.33)	78.68	(16.88)
Boys	79.02	(16.79)	78.47	(16.63)	78.91	(16.90)	79.22	(19.28)	78.82	(17.10)
Difference Girls - Boys ($\Delta_2 = y_1^0 - y_0^0$)	-1.35	(1.55)	-0.48	(1.80)	0.20	(2.22)	-0.31	(3.73)	-0.14	(1.53)
Number of teachers	33		33		33		30		33	
Number of pupils	496		348		232		102		496	
Differences $\Delta_1 - \Delta_2$	-1.62***	(0.61)	-1.14**	(0.52)	-1.52***	(0.47)	-0.40	(0.54)	-0.61**	(0.27)

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level. S.E. for differences.

5 Results

Within class between child gender estimation model

In Table 4, the results for mathematics are described. In this analysis, the difference between matched girls and boys in standardized mathematics posttest score is the dependent variable. In the first column we display the results for all matched couples. Then, in Columns 2 and 3, the sample is narrowed down to the best 75% and 50% matches. In Column 4, the results for the analysis based on matching with replacement[6] are displayed. For interpretation purposes, we generated dummies for the control variables, which where originally also differences between girls and boys.

With the full sample with all available matches, we have found a significant negative coefficient. When having an opposite-gender teacher, the difference between the posttest scores of girls and boys changes with 11.2% of a standard deviation, which equals a change in score of approximately 1.5 points. This can either mean that girls perform better with a male teacher, that boys perform better with a female teacher, or even both. In the next section, we will examine whether this opposite-gender teacher effect occurs for girls with a male teacher, for boys with a female teacher, or for both. In the model with the best 75% and 50% matches, the significant effect of the teachers' gender has disappeared, with non-significant estimates for the TFEM dummy. When we have estimated our model with sub-samples of better girl-boy matches, the results no longer suggest the existence of the opposite-gender effect found in the estimation with the full sample. We also performed the matching with replacement. When the girl-boy couples are very good matches, based on pretest score, the coefficient for the TFEM dummy is highly insignificant. Apparently with better matches, we do not find any indication of the existence of a same-gender, or opposite-gender teacher effect.

[6] In case of classes with a majority of girls, one boy could be matched to multiple girls if he is the best match for those girls. In case of classes with a majority of boys, one girl could be matched to multiple boys if she is the best match for those boys.

Table 4: Results: within class between child first differences model, for all matches, best 75%, best 50% and for matching with replacement

	All	75%	50%	Replacement
Teacher gender - Female	-0.112**	-0.014	-0.074	-0.053
	(0.052)	(0.056)	(0.066)	(0.043)
Ethnicity: (reference: both Dutch)				
Girl non-Dutch background	-0.066	0.059	0.005	-0.049
	(0.067)	(0.075)	(0.095)	(0.056)
Boy non-Dutch background	0.091	0.108	-0.020	0.027
	(0.066)	(0.073)	(0.084)	(0.054)
Both non-Dutch background	0.043	0.080	-0.050	0.095**
	(0.054)	(0.057)	(0.068)	(0.045)
Education level mother: (reference: both low)				
Girl low - boy middle	-0.139	-0.042	0.127	-0.070
	(0.124)	(0.136)	(0.166)	(0.105)
Girl middle - boy low	-0.068	0.013	0.039	-0.022
	(0.125)	(0.138)	(0.165)	(0.103)
Both middle	-0.011	0.042	0.107	-0.005
	(0.103)	(0.112)	(0.136)	(0.083)
Girl middle - boy high	0.085	0.125	0.195	0.067
	(0.129)	(0.138)	(0.163)	(0.103)
Girl high - boy middle	0.172	0.260*	0.266	0.088
	(0.128)	(0.139)	(0.174)	(0.108)
Both high	0.076	0.134	0.105	-0.008
	(0.114)	(0.125)	(0.152)	(0.092)
Girl low - boy high	0.099	0.120	0.237	0.063
	(0.151)	(0.169)	(0.193)	(0.136)
Girl high - boy low	0.016	-0.033	0.123	-0.273**
	(0.143)	(0.156)	(0.189)	(0.120)
Girl and/or boy unknown	0.095	0.126	0.270*	0.080
	(0.113)	(0.123)	(0.149)	(0.092)
Age in days * 100, in differences	0.053***	0.039***	0.038***	0.028***
	(0.008)	(0.009)	(0.010)	(0.007)
Constant	0.104	0.027	0.094	0.148*
	(0.111)	(0.121)	(0.145)	(0.090)
Adjusted R^2	0.043	0.021	0.013	0.020
Sample	1293	902	597	1293

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level.

6 Robustness checks

In this section, we address several issues that may bias the teacher-gender estimate. First of all, it might be that male teachers self-select themselves in higher, or lower, grades. If potential learning gains are grade-dependent, then this may impose a bias on our teacher-gender estimate. Second, a non-random assignment of teachers to classes could also bias the results. The estimated same-gender parameter could be biased, because teachers could self-select themselves either in strong or weak classes. If both male, and female teachers select themselves to a similar extent in strong and weak classes, this is not a problem. If this is not the case however, $\gamma_2$can be biased because unobserved differences in background characteristics may arise, because children are selectively distributed over classes and schools. Third, up until now we do not take potential effects of previous class characteristics into account. Perhaps, there could be a same-gender or opposite-gender effect, but it already had an influence in the previous year. Possibly, an effect of teacher gender effect would be more likely to benefit younger children. Furthermore, our teacher-gender estimate might also be biased, because of differences in the level of math in the previous class. Therefore, we include previous class teacher gender, and average class pretest score based on the previous year class composition. Finally, we estimate whether equation 4 is equal to equation 3, i.e. we estimate our model with additional school, class, and teacher characteristics, which should not have an impact on our estimates, since these are constant for the matched girl-boy couples. This section starts with a separate analysis by grade. Second, we consider to what extent the assignment of teachers to classes affects our results. Third, we examine whether the teacher gender and average class math score of the previous school year affect the differences in posttest score between the girl-boy couples in the current school year. Finally, as mentioned in 3, as a further robustness check, we also estimate 3.

Grade analysis

Table 5 shows the math pre-testscores of boys and girls, separately for each grade, and for male and female teachers. The percentage of male teachers shows that male teachers teach grade 5 classes more frequently, compared to grade 3 and grade 4 classes. It might be that male teachers select themselves more often into higher grades, and/or female teachers select themselves more often into lower grades. In Table 5 we compare the math pretest scores of boys and girls again for male and female teachers, separate per grade.

Both in grade 3, and in grade 5 classes with male teachers, boys and girls have similar pretest scores in mathematics. Only in grade 4 classes with male teachers, boys seem to perform better, but the difference is not significant. Contrary to this, boys score significantly higher than girls in all grades with female teachers. If we compare the test scores between female and male teachers, we find pupils of male teachers scoring higher in grade 3, but pupils of female teachers scoring higher in grade 4. For grade 5, both the difference between boys and girls, and the difference between male and female teachers is very small. Male teachers teach more often grade 5 classes, in comparison to grade 3, and grade 4 classes. It is possible that the potential growth differences are not constant for boys and girls in different grades. To account for differences between to what extent male and female are likely to teach different grades, we performed the analysis separately for grade 3, 4, and 5. In Table 6 the results appear.

If we split the samples based on grades, we only find a significant estimate for grade 3 classes. It appears that the significantly negative coefficient for the full sample was mainly driven by the grade 3 classes.

Table 5: Average mathematics pretest scores for boys and girls with male and female teachers, per grade

	Grade 3		Grade 4		Grade 5	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Female Teachers						
Girls	60.29	(14.69)	76.64	(14.09)	88.24	(12.58)
Boys	63.48	(15.29)	79.73	(13.27)	90.52	(12.60)
Difference Girls - Boys ($\Delta_1 = y_1^1 - y_0^1$)	-3.19***	(1.23)	-3.08***	(1.03)	-2.28**	(0.94)
Number of classes	46		51		52	
Number of pupils	592		792		724	
Male Teachers						
Girls	66.35	(14.93)	72.46	(17.04)	89.49	(11.83)
Boys	65.80	(14.29)	76.48	(14.09)	90.33	(11.39)
Difference Girls - Boys ($\Delta_2 = y_1^0 - y_0^0$)	0.55	(2.55)	-4.01	(2.62)	-0.84	(1.65)
Number of classes	8		11		13	
Number of pupils	132		142		198	
% male teachers in grade	18.2		17.0		21.5	
Differences $\Delta_1 - \Delta_2$	-3.73***	(0.77)	0.93	(1.50)	-1.44*	(0.86)

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level. S.E. for differences.

Table 6: Analysis separate per grade

	Grade 3	Grade 4	Grade 5
Teacher gender - Female	-0.271**	0.081	-0.051
	(0.105)	(0.092)	(0.085)
Pupil characteristics (in differences)	yes	yes	yes
Adjusted R^2	0.072	0.025	0.051
Sample	362	417	461

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level.

Table 7: Average mathematics posttest scores for boys and girls with male and female teachers, per grade

	All		Grade 3		Grade 4		Grade 5	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Female Teachers								
Girls	87.78	(16.82)	74.83	(14.42)	88.64	(12.32)	98.90	(12.90)
Boys	91.09	(16.57)	79.04	(14.48)	90.82	(13.77)	101.64	(13.14)
Difference Girls - Boys ($\Delta_1 = y_1^1 - y_0^1$)	-3.30***	(0.73)	-4.21***	(1.19)	-2.18**	(0.99)	-2.74***	(0.97)
Male Teachers								
Girls	91.02	(15.82)	82.77	(13.94)	85.94	(14.04)	100.74	(12.37)
Boys	92.34	(15.38)	82.12	(13.21)	89.97	(11.46)	101.38	(13.68)
Difference Girls - Boys ($\Delta_2 = y_1^0 - y_0^0$)	-1.33	(1.40)	0.65	(2.36)	-4.03*	(1.03)	-0.65	(1.85)
Differences $\Delta_1 - \Delta_2$	-1.97**	(0.80)	-4.86***	(1.50)	1.85	(1.58)	-2.09*	(1.23)

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level. S.E. for differences.

Separately for grade 3, we find a larger negative coefficient then for the full sample. For the pupils from grade 4 and grade 5 classes, there is no evidence for the existence of either a same-gender, or an opposite-gender teacher effect. To investigate whether the negative estimate for grade 3 is being caused by girls performing better when being taught by a male teacher, boys performing better when being taught by a female teacher, or both, in Table 7 the average mathematics posttest scores per grade are shown.

The posttest scores differ substantially between female teachers and male teachers. While boys perform significantly better than girls in the math posttest score in classes with a female teachers, when being taught by male teachers, there are on average no differences in posttest scores between boys and girls. Especially in grade 3, there is a large gap between the children taught by male teachers, and the children taught by female teachers. While the difference in test scores for boys is approximately 3 points, for girls it is almost 8 points. When being taught by female teachers, boys perform better than girls in all grades. In classes with male teachers on the other hand, boys and girls have equal posttest scores. Only in grade 4 boys have significantly higher posttest scores. Apparently, in grade 3, girls perform better when being taught by a male teacher. Rather than an opposite-gender teacher effect, this appears to be caused by girl-boy couples which do not have equal pretest scores.

Do male teachers select themselves in better classes?

Until now, we have not discussed that the estimated same-gender parameter can be biased, because teachers could select themselves either in strong, or weak classes. If both male and female teachers select themselves to a similar extent in strong and weak classes, this is not a problem. If this is not the case however, γ_2 can be biased, since unobserved differences in background characteristics may arise, because children are selectively distributed over classes and schools. It is worth mentioning that, in the Dutch education system (unlike many other systems, such as the US) parents choose to which school their child goes[7]. Therefore, the non-random distribution of ability over schools is, to a large extent, driven by parental choices. A major advantage of the within classes between child gender approach, is that it compares girls and boys within classes such that it accounts for much of the bias that arises due to the non-random distribution of children to schools. This does not, however, guarantee that teachers are randomly distributed over primary schools. Even though it is unlikely that male teachers are structurally better teachers than female teachers (or vice versa), it may happen that, for example, female teachers select themselves in weak classes. This may potentially affect the estimation results, because the variable TFEM could then pick up the effect that learning growth differences for weak and strong classes are not the same, which is obviously not a same-gender teacher effect. We should therefore verify how teachers are distributed over schools, and compare how the past performance of children of male teachers relate to that of female teachers.

In the extended school time survey of 2010, we asked school directors whether their school has parallel classes[8], and how schools assign their teachers to these parallel classes. The answers to these questions enable us to assess whether our results could be affected by the way how schools assign teachers to their different (parallel) classes. We assume that when schools do not have parallel classes, the pupils within that school are assigned randomly to their classes, since they are only assigned based on their grade, i.e. age. If a school divides their pupils of the same grade in a low-achieving and in a high-achieving group, then we can no longer speak of random assignment of pupils. If male and female teachers are randomly assigned to these stronger and weaker parallel classes, then this does not have to affect the same-gender/opposite-gender effects, if these effects are homogenous, i.e. not heterogeneous for different intelligence levels. It is however possible that better teachers are selectively assigned to better or worse classes over grades, when schools have, for instance, a policy to, (1) let their best pupils excel, or (2) focus on reducing the deprivation of their weakest pupils. We investigate whether controlling for non-random assignment of teachers to classes will change the TFEM estimate reported in Table 4. We estimate the same model as before, but include an interaction term of the TFEM dummy with a dummy, which equals 1 when teachers are randomly assigned to parallel classes, and equals 0 when teachers are not randomly assigned to parallel classes, or when this information is unknown. We have this information for 781 of the 1293 girl-boy couples. Because we do not know the assignment policy for every school, we also include an interaction term of the TFEM dummy with a dummy, which equals 1 if the assignment policy is unknown, and 0 if teacher assignment is either random or non-random. The estimation results are shown in Table 8.

The estimates for the teacher gender - female dummy are similar to our results reported in Table 4. We find a significantly negative estimate for the full sample, but insignificant estimates once we limit the sample to acquire a better match in pretest scores. In addition, both random assignment dummy interaction terms are insignificant. This suggests that non-random assignment of teachers to classes does not influence our estimates.

[7]Although some schools do have a policy which enables them to give priority to children who live within certain postal code regions.

[8]Parallel classes implies in this case more than one class of the same grade

Table 8: Analysis controlled for teacher assignment policy

	All	75%	50%
Teacher gender - Female	-0.120**	0.001	-0.046
	(0.058)	(0.062)	(0.074)
Teacher gender - Female * Teachers random	0.007	-0.022	-0.043
	(0.065)	(0.071)	(0.084)
Teacher gender - Female * Teachers random unknown	0.017	-0.030	-0.053
	(0.050)	(0.054)	(0.064)
Pupil characteristics (in differences)	yes	yes	yes
Adjusted R^2	0.041	0.020	0.011
Sample	1293	902	597

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level. Included pupil characteristics are the same as in Table 4, but not shown.

Is the same-gender teacher estimate influenced by previous class characteristics?

Previous class characteristics, like the teacher's gender of the previous school year class of the pupils, or the average math level of this previous class, could also bias our estimates. If the gender of the current teacher matters for the math performance of pupils, then the previous teacher gender should matter as well. Some pupils had an opposite-gender teacher in their previous school year, and a same-gender teacher in their current school year. Others had a same-gender teacher in their previous school year and an opposite-gender teacher in their current school year. There are also pupils who had 2 years in a row a same-gender or opposite-gender teacher. Note that girls, not only have a higher probability to be in a class with a same-gender teacher in their current school year, but also in their previous school year. Our main analysis is based on the school year 2010-2011, so in this case the teacher's gender of their 2009-2010 class, one grade lower. To examine the influence of the pupils' last year teacher gender, we estimate the within class between child gender estimation model with an additional teacher gender (female) dummy for the previous school year. We also include a dummy, which equals 1 if the teacher gender of the previous school year is unknown, and 0 otherwise. We have this information for 677 of the 1293 girl-boy couples.

Not only the teacher's gender of the previous school year could matter, it could also be that pupils in the current school year benefit more (or less) from having a same-gender teacher if their class from the previous school year had on average a high (or low) level of mathematics proficiency. Being in a class with other pupils with high math scores could be related to different opportunities to improve ones mathematics level in the present school year. Therefore, we also include the average math pretest score, which is the test pupils have taken at the end of the school year 2009-2010. We cannot simply use the per class average of the pretest score, because not all classes remained the same in pupil composition. Therefore, we calculated the average over the class identifier of the 2009-2010 school year. To be able to utilize our full sample, we imputed the missing values for average class pretest score with the general average for this variable, and included a dummy which equals 1, if the value for average class pretest score was imputed, and 0 otherwise. This was the case for 233 of the 1293 girl-boy couples. In addition, we included a dummy variable which equals 1 if parallel classes are changed every school year, and equals 0, if these classes remain the same every

Table 9: Analysis with controls for teacher gender and average class test score of the previous school year

	100%	75%	50%
Teacher gender - Female 2010-2011	-0.131**	-0.029	-0.095
	(0.052)	(0.056)	(0.067)
Teacher gender - Female 2009-2010	-0.031	-0.050	-0.005
	(0.070)	(0.078)	(0.092)
Teacher gender 2009-2010 unknown	0.070	0.058	0.098
	(0.075)	(0.084)	(0.100)
Average class pretest score (end of 2009-2010)	0.002	0.002	0.004
	(0.002)	(0.002)	(0.003)
Average class pretest score (end of 2009-2010) unknown	-0.160**	-0.172**	-0.069
	(0.081)	(0.087)	(0.103)
Class composition changed (reference: no change)	0.114*	0.116	0.008
	(0.069)	(0.075)	(0.090)
Class composition change unknown	0.004	0.043	-0.093
	(0.064)	(0.069)	(0.082)
Pupil characteristics	yes	yes	yes
Adjusted R^2	0.045	0.023	0.014
Sample	1293	902	597

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level.

school year, or when this information is unknown. We also include a dummy which equals 1, if changes in class composition is unknown and 0 otherwise. We have this information for 962 of the 1293 girl-boy couples. Table 9 shows the estimation results for our within class between child gender estimation model, with additional regressors for previous school year teacher gender, and average class math test score of the previous class.

The results show that the TFEM coefficient is hardly influenced by the inclusion of the variables about the previous class average math score and teacher gender. For every model, the TFEM coefficient is similar to the TFEM coefficient of the main analysis, although slightly larger in size. The teacher gender of the previous school year dummy is highly insignificant, just like the dummy indicating whether previous year teacher gender is unknown. The average class pretest score is also not related to the girl-boy differences in posttest scores.

Do school, class and teacher characteristics still matter in the within class between child gender estimation model?

In Section 3, we explained how in the within class girl-boy couples the variation on the school, class, and teacher level should be eliminated from the estimation model. The math performance of girls and boys within couples is likely influenced in a similar way by school, class and teacher characteristics, which implies that the girl-boy differences for these indicators equal 0. Perhaps, boys and girls are, in fact, differently

Table 10: Analysis with additional school, class and teacher characteristics

	100%			75%			50%		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Teacher gender - Female	-0.124**	-0.125**	-0.141***	-0.022	-0.023	-0.040	-0.075	-0.074	-0.095
	(0.052)	(0.052)	(0.053)	(0.057)	(0.057)	(0.057)	(0.068)	(0.068)	(0.068)
School characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Class characteristics	no	yes	yes	no	yes	yes	no	yes	yes
Teacher characteristics	no	no	yes	no	no	yes	no	no	yes
Adjusted R^2	0.044	0.044	0.047	0.021	0.019	0.025	0.010	0.007	0.014
Sample	1293	1293	1293	902	902	902	597	597	597

Note: * = significant at the .10 level. ** = significant at the .05 level. *** = significant at the .01 level. All models estimated with pupil characteristics as used in Table 4.

influenced by certain constant school, class, or teacher characteristics, for instance when boys' and girls' math performance is differently influenced by a high level of disadvantaged students at school. To check whether equation 3 and equation 4 are indeed equal, we estimated equation 3, by including several school, class, and teacher characteristics. The school characteristics are: total pupils in school, a dummy which equals 1, if the school is in a poverty problem area, the percentage of children from parents with very low, or no education, the percentage of children from parents with low education, and the percentage of boys in school. The class characteristics are: the percentage of boys in the class, and class size (amount of pupils). We only use total work experience as teacher characteristic, because there is hardly any variation in the education level of the teachers in our sample. Table 10 shows the results for our estimation model with additional school, class, and teacher characteristics.

We have stepwise included school characterics (1), class characteristics (2) and teacher characteristics (3). The Inclusion of school, and class characteristics does hardly alter our estimate for the teacher gender - female dummy with the full sample. If we, however, include teacher experience in the model, the estimate increases slightly to 14.1% of a standard deviation. The estimation models with only the best 75% and the best 50% matches yield similar results as in the main analysis. The TFEM dummy remains not significantly different from zero. We can conclude that the results of estimating equation 3 and equation 4 are very similar. School, class and other teacher characteristics, besides gender do not influence our TFEM dummy estimate.

7 Concluding remarks

Traditionally, there have been gender differences in school performance. While boys in general performed better in mathematics and science, girls performed better in language and reading. These achievement differences have long been considered as given, but the diminishing (and sometimes reversed) achievement gaps, together with the observed higher share of girls in higher secondary education levels and of women among first-year students in Higher Education, have changed the traditional expectations about gender differences in education. At the same time, primary school teaching has become, to a high extent, a female profession, which has led to a lack of male role models in primary education. This could negatively influence the school achievement of boys. We investigate whether there exists a same-gender teacher effect: girls performing better with a female teacher and boys performing better with a male teacher. Using a within class between child gender approach, we identify the effect of having a same-gender / opposite gender teacher on mathematics performance of primary school pupils. In this approach we construct girl-boy couples within classes, based on having a similar math score at the end of the previous school year. We estimate the effect of having a same-gender / opposite gender teacher on the difference between within couple girl and boy test scores at the end of the current school year, while eliminating all school, class and teacher effects, with the exception of teacher's gender. We neither find evidence for the existence of a same-gender teacher, nor of an opposite-gender teacher effect for mathematics. We do find an opposite-gender teacher effect if we estimate our model for the full sample, but these findings are not reliable, because the matched girls and boys within classes still have significantly different pretest scores. If we only consider the couples with the best 75% or 50% matches per class, the significant differences between the matched boys and girls disappear. When we restrict the sample in our estimation model to the best 75% and the best 50% matches, we no longer find any significant effect of having a same-gender or opposite-gender teacher. When we estimate a separate analysis per grade, we find that the significant negative coefficient for the full sample is driven by grade 3 classes. In grade 3, girls with male teachers perform significantly better in math, then girls with female teachers. Rather than an opposite-gender teacher effect, this appears to be caused by girl-boy couples which do not have equal pretest scores. Our non-significant results do not change if we correct for whether teachers are randomly assigned to classes or not. The way how teachers are assigned to their classes could not have biased our results. After we have investigated the potential ongoing effects of the teachers' gender and the average math level in the class of the previous school year, we can conclude that both the previous teacher gender and the average class mathematics level did not affect our estimates. Finally, also the inclusion of school, class and teacher characteristics does not change our results. We conclude that there has neither been a same-gender, nor an opposite-gender teacher effect for mathematics. Differences in math performance of boys and girls do not seem to be related to their teacher's gender. Efforts to get more men into teaching in primary education will most likely not influence gender differences in math performance.

References

[1] Alexander, K.L. and Eckland, B.K. (1974) Sex differences in the educational attainment process. Am Sociol Rev 39(5):668-682

[2] Alexander, K.L., Entwisle, D.R., and Dauber, S.L. (2003) On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades. Cambridge, UK: Cambridge Univ Press

[3] Alexander, K.L. and McDill, E. (1976) Selection and allocation within schools: some causes and consequences of curriculum placement. Am Sociol Rev 41(6):963-980

[4] Ammermüller, A. and Dolton, P. (2006) Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and in the USA. ZEW Discussion Paper No. 06-060

[5] Buchmann, C., DiPrete, T.A., and McDaniel, A. (2008) Gender Inequalities in Education. Annu Rev Sociol 34:319-337

[6] Coenen, J.B., Meng, C. and Van der Velden, R. (2011) Schoolsucces van jongens en meisjes in het HAVO en VWO: waarom meisjes het beter doen. ROA-R-2011/2. ROA, Maastricht

[7] Corcoran, S.P., Evans, W.N. and Schwab, R.M. (2004) Women, the Labor Market, and the Declining Relative Quality of Teachers. J Pol Anal Manag 23(3):449-470

[8] Dee. T.S. (2007) Teachers and the gender gaps in student achievement. J Hum Resource 42:528-554

[9] Driessen, G. (2007) The feminization of primary education: effects of teachers' sex on pupil achievement, attitudes and behaviour. Int Rev Educ 53:183-203

[10] Duckworth, A.L. and Seligman, M.E.P. (2006) Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. J Educ Psychol 98(1):198-208

[11] Entwisle, D.R., Alexander, K.L., Olson, L.S. (2007) Early schooling: the handicap of being poor and male. Sociol Educ 80(2):114-138

[12] Falch, T. and Naper, L.M. (2011) Educational Evaluation Schemes and Gender Gaps in Student Achievement. WORKING PAPER SERIES 04/2011 Department of Economics, Norwegian University of Science and Technology (NTNU)

[13] Feingold, A. (1988) Cognitive gender differences are disappearing. Am Psychol 43(2):95-103

[14] Graue, M.E. and DiPerna, J. (2000) Redshirting and early retention: who gets the "gift of time"and what are its outcomes? Am Educ Res J 37(2):509-534

[15] Hedges, L.V. and Nowell, A. (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 269(5220):41-45

[16] Hinshaw, S.P. (1992) Externalizing behaviour problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. Psychol Bull 111(1):127-155

[17] Holmlund, H. and Sund, K. (2008) Is the gender gap in school performance affected by the sex of the teacher? Lab Econ 15:37-53

[18] Hyde, J.S., Fennema, E., Ryan, M., Frost, L.A. and Hopp, C. (1990) Gender comparisons of mathematics attitudes and affect: a meta-analysis. Psychol Women Q 14(3):299-342

[19] Klaveren, van C. (2011) Lecturing style teaching and student performance. Econ Educ Rev 30:729-739

[20] Leahey, E., and Guo, G. (2001) Gender differences in mathematical trajectories. Soc Forces 80(2):713-732

[21] Maccoby, E.E. and Jacklin, C.N. (1974) The Psychology of Sex Differences. Palo Alto, US: Stanford University Press

[22] Malone, L.M., West, J., Denton, K.F. and Park, J. (2006) The Early Reading and Mathematics Achievement of Children Who Repeated Kindergarten or Who Began School a Year Late. Washington, DC: National Center for Education Statistics

[23] Marks, G.N. (2007) Accounting for the gender gap in reading and mathematics: evidence from 31 countries. Oxf Rev Educ 34(1):1-21

[24] Mechtenberg, L. (2009) Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievement, Career Choices and Wages, Rev Econ Stud 76:1431-1459

[25] Merrell, C. and Tymms, P. (2011) Changes in Children's Cognitive Development at the Start of School in England 2001– 2008. Oxf Rev Educ 37(3):333-345

[26] Meyer, E. and Van Klaveren, C. (2011) Evaluation of an Extended Day Program in the Netherlands: A Randomized Controlled Field Experiment, TIER Working Paper Series

[27] Mickelson, R.A. (1989) Why does Jane read and write so well? The anomaly of women's achievement. Sociol Educ 62(1):47-63

[28] Neugebauer, M., Helbig, M. and Landmann, A. (2010) Unmasking the Myth of the Same-Sex Teacher Advantage. Eur Socio Rev 27(5):669-689

[29] Ouazad, A. (2008) Assessed by a Teacher Like Me: Race, Gender, and Subjective Evaluations. INSEAD Working paper 2008/57/EPS

[30] OECD (2010) PISA 2009 at a Glance, OECD Publishing. http://dx.doi.org/10.1787/9789264095298-en

[31] Perkins, R., Kleiner, B., Roey, S. and Brown, J. (2004) The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000. Washington, DC: National Center for Education Statistics

[32] Pollack, W. S. (2006) The 'war' for boys: hearing 'real boys' voices, healing their pain. Prof Psychol Res Pract 37:190-195

[33] Sexton, P. (1969) The Feminized Male: Classroom White Collars, and the Decline of Manliness, New York Random House

[34] Sokal, L., Katz, H., Chaszewski, L. and Wojcik, C. (2007) Good-bye, Mr. Chips: Male Teacher Sortages and Boys' Reading Achievement. Sex Roles 56:651-659

[35] Willingham, W.W. and Cole, S.E. (1997) Gender and Fair Assessment. Mahwah, US: Lawrence Erlbaum

[36] Yatchew, A. (1998) Nonparametric regression techniques in economics. J Econ Lit 36(2):669-721