Is the Gender Gap in School Performance Affected by the Sex of the Teacher?*

Helena Holmlund†
Krister Sund‡
Swedish Institute for Social Research (SOFI)
Stockholm University

November 4, 2005

Abstract
Girls outperform boys in school. We investigate whether the gender performance gap can be attributed to the fact that the teacher profession is female dominated, that is, is there a causal effect on student outcomes from having a same-sex teacher? Using data on upper-secondary school students and their teachers from the municipality of Stockholm, Sweden, we find that the gender performance differential is larger in subjects where the share of female teachers is higher. We argue, however, that this effect can not be interpreted as causal, mainly due to teacher selection into different subjects and non-random student-teacher matching. Exploring the fact that teacher turnover and student mobility give rise to variation in teacher's gender *within student and subject*, we estimate the effect on student outcomes of changing to a teacher of the same sex. We find no strong support for our initial hypothesis that a same-sex teacher improves student outcomes.

---

* The authors wish to thank Anders Björklund, Mikael Lindahl, Marianne Sundström and seminar participants at the Princeton University Labor Lunch, the Swedish Institute for Social Research and the CEPR conference on "Economics of Education and Education Policy" in Uppsala, October 2005, for valuable comments. Financial support from Jan Wallanders and Tom Hedelius foundation and The Swedish Council for Working Life and Social Research is gratefully acknowledged.
† Swedish Institute for Social Research (SOFI), Stockholm University, SE-106 91 Stockholm, SWEDEN. E-mail: Helena.Holmlund@sofi.su.se, phone: +46 (0)8 16 22 30.
‡ Swedish Institute for Social Research (SOFI), Stockholm University, SE-106 91 Stockholm, SWEDEN. E-mail: Krister.Sund@sofi.su.se, phone: +46 (0)8 16 23 07.

I. Introduction

Girls outperform boys in school. While it is well known that girls score significantly higher than boys on for example reading tests, there is now increasing evidence that the gender gap in school performance is closing in math and science, subjects thought of as being dominated by boys. For example, U.S. educational statistics report that between 1973 and 1999, the male advantage in mathematics and science scores at age 17 was significantly reduced (Campbell, Hombo and Mazzeo 1999). Further evidence from a different country is test scores at age 15 in Sweden. While girls clearly score higher than boys on Swedish and English tests, there is no obvious gender difference in mathematics (Swedish National Agency for Education 2004).

In this paper we take the increasing feminization of the teacher profession as a starting point for studying the gender performance differential. Of all Swedish upper-secondary school teachers, 41 percent were female in 1973, compared to 48 percent in 2003 (Statistics Sweden 1994, 2001, 2004). In particular, we explore whether the gender performance gap can be explained by the gender of the teacher, that is, does the girls' performance lead, measured in terms of grades, increase with a female teacher? There are a number of possible hypotheses that can explain such a finding. First, teachers might have preferences over students of their own sex, and hence female (male) teachers will, given student performance, reward girls (boys) more highly in terms of grades. Second, if not preferences, gender stereotypes may influence teachers' evaluation of their students. Both of the aforementioned hypotheses we label as discrimination. Third, we may think of teachers as role models for the students. If students identify themselves more with same-sex role models, it is possible that performance will be enhanced when students have a teacher of their own gender. Fourth, a result where same-sex teachers improve student outcomes is also consistent with the theory of stereotype threat (Steele 1997). This theory states that in the case of negative stereotypes against a group, group members may internalize the stereotypes as explanations to their own behaviour. These two last hypotheses we jointly label as the role model explanation.

From the economist's point of view, the gender gap is a concern for several reasons. If discrimination in grades is prevalent, discrimination-based access to higher education will introduce an efficiency loss; on the margin, a more able student will be denied entry to university education while a less able but higher graded student will enter instead (assuming that teacher's assessment of the student, in terms of grades, is a screening tool for access to higher education, which is true in the case of Sweden). On the other hand, if role models are important for student outcomes and enter the education production function, we should pay attention to this in formulating a policy that aim at maximizing output.

In this paper, we study student outcomes in terms of course grades in the Swedish upper-secondary school. For each course the student takes, we have matched information on both the gender of the student and the teacher. Our main interest is to estimate a causal effect of having a same-sex teacher on student outcomes. The empirical challenge is to find an identification strategy that allows us to interpret our results as causal, since teachers are not randomly assigned to students and subjects.

We use two different identification strategies to study the effect of teacher's gender on student outcomes. First, we assume that teacher's gender is exogenous and run an OLS regression. Second, we estimate an equation of student achievement growth, exploring variation in teacher's gender within a specific subject, introduced by a change of teacher during the three year course of upper-secondary school. This within-student identification strategy controls for unobserved student characteristics, teacher sorting into different subjects, and also captures the influences of past school and teacher characteristics on current achievement. The data do not allow us to separate between discrimination and role models as being the primary explanation for why a same-sex teacher should have a positive effect on student outcomes. However, since the previous evidence is relatively scarce, we think that a first step of establishing whether there is an effect is in itself a contribution to the literature.

Our findings show that the female-male performance gap is clearly higher in subjects where the teachers are predominately female. Thus, an OLS regression supports our initial hypothesis that a same-sex teacher is positively associated with student grades. However, we argue that this finding cannot be interpreted as a causal effect; teachers are not allocated randomly across subjects, and it is likely that female teachers sort into subjects where they themselves (and also their female students) perform the best, regardless of any student-teacher interactions. Our baseline OLS result is therefore mainly an effect of a spurious correlation between teacher's gender and student performance, introduced by gender based sorting of teachers into different subjects. Moreover, the OLS estimate may be biased due to correlations between current and past teacher characteristics, and student-teacher matching might be non-random. To address the causality issue, we use variation in teacher's gender over time, introduced by teacher or student mobility, to estimate the effect within student and within subject. The results of this empirical strategy are meant to bring an answer to the relevant policy question "What would happen to the gender gap in school performance if the teacher profession were more balanced in terms of the sex of the teacher?". Once we control for unobserved student characteristics, the influence of past teacher characteristics and teacher sorting, we find no strong support for the hypothesis that a same-sex teacher improves student outcomes.

The remainder of the paper has the following structure: Section II discusses the related literature, section III gives an overview of the institutional setting, section IV describes the data, the econometric approach and reports on the findings, and section V concludes.

II. Related Literature

Related empirical evidence of the impact of matched teacher-student characteristics looks at both race and sex. Until recently, the bulk of the literature focusing on gender has directed its interest towards tertiary education. Lately however, Thomas S. Dee (2005b) studies the effect of having a teacher of the same sex in 8^{th} grade in US middle schools. Using within-student variation across subjects he finds that having a same-sex teacher has substantial positive impacts, both on test scores, student interest in the subject, and teacher assessments of students. Another recent example is Lavy (2004), who studies the importance of gender stereotypes in the evaluation of student outcomes at upper-secondary level in Israel. Using a natural experiment which allows for comparisons between a gender-blind test score and a non-gender-blind score, he finds, contrary to his expectations, that the gender bias is in favour of girls. Girls have systematically higher scores on the non-blind test compared to the blind test, but there seems to be a small systematic difference based on teacher's gender in the discrimination against boys. The gender bias in favour of girls is larger among male math teachers than female math teachers, but the opposite is true for physics, biology and chemistry.

Earlier studies on gender interactions have directed its interest towards tertiary education, and the task has been to investigate whether female role models affect educational outcomes. Rothstein (1995) finds a positive association between female students post graduate education and the share of female faculty. This finding is in line with a role model mechanism, but the empirical strategy cannot rule out another potential explanation: female faculty and students might self-select into academic environments that are supportive of women. Neumark and Gardecki (1998) study role model and mentoring effects in U.S. Economics Ph.D. programs. They find no support for the hypothesis that a larger number of female faculty or a female dissertation chair positively affects future success for the female students. More recently, Bettinger and Long (2005) assess whether the gender of the instructor in an introductory college course affects subsequent choice of courses in the particular subject. The evidence is somewhat mixed, but Bettinger and Long conclude that their results, being particularly strong for mathematics and statistics, geology, sociology and journalism, support the role model hypothesis.

Further studies that discuss role models in academia are Canes and Rosen (1995), Solnick (1995), Dynan and Rouse (1997), Ashworth and Evans (2001) and Rask and Bailey (2002).

Most studies of primary and secondary education have had race as the main focus. Ehrenberg and Brewer (1995) revisit the Coleman Report data, and study how gain scores of synthetic cohorts are affected by teachers' verbal ability and race. They find that black teachers increase the gain scores of black students at secondary level, while decreasing the gain scores of white students at both primary and secondary levels. Ehrenberg, Goldhaber and Brewer (1995) analyze how teachers' race, gender and ethnicity influence both student outcomes and the subjective teacher evaluation of students, depending on the race, gender and ethnicity of the student. They find no evidence supporting the hypothesis that a same-sex or same-race teacher improves student test scores. However, they do find that matched teacher-student characteristics in terms of sex and race have an impact on the teacher's subjective assessment of the students. More recently, Dee (2004) uses the experimental design of the Tennessee's Project STAR data in which students and teachers were randomly assigned to different classes. The results show that a same-race teacher improves student outcomes, both for black and white students. In Dee (2005a) there is further evidence that assignment to a teacher with the same race, ethnicity or gender has significant impacts on teacher assessments of student behaviour.

III. The Swedish Upper-Secondary School

The Swedish educational system requires 9 years of compulsory schooling. After compulsory school, municipalities offer a three year upper-secondary school.[1] Entry into upper-secondary school is based on the GPA from compulsory school graduation.[2] Almost 98 percent of the graduates continue to upper-secondary education, either within an academic track (with specialization in math/science or humanities/social science) or within one of several vocational branches (examples are construction, hairdressing and chef education). In the school year 2002/03, 42 percent of the students who continued to upper-secondary education enrolled in an academic track (Statistics Sweden 2004).[3] The vocational tracks include a minimum of theoretical subjects, granting eligibility to some higher education.

The upper-secondary school is structured in subjects and courses. Subjects are for example mathematics, Swedish and English, while within each subject the student takes several courses (for example math A, math B, math C etc). Each course is evaluated separately and the student is given a final course grade. Past school performance should not affect the specific course grade; only the actual performance in the course, and national test results, are taken into account when the grade is set.[4] The grading system of

---

[1] The Swedish municipalities are required by law to furnish eligible students with upper-secondary education.

[2] Students are required to have graduated from compulsory school with at least pass in English, math and Swedish (or Swedish as a second language). Students are then ranked according to a GPA that is based on the best 16 subjects.

[3] The two academic tracks are the natural science programme and the social science programme. In the school year of 2000/01, however, the technology programme was introduced. (The technology programme was previously incorporated in the natural science programme, but by 200/01 it was introduced as an independent track).

[4] We emphasise that the courses are graded separately since we will use teacher turnover in one of our identification strategies. If previous performance were taken into account, the new and the old teacher assessment of the student would be correlated. It is hence important to note that a new teacher should *not* consider previous performance.

Swedish upper-secondary school is a criterion-referenced grade scale. Grades are scaled in four levels: fail, pass, pass with distinction and pass with special distinction. Grade setting in Swedish schools is highly decentralized, meaning that schools and teachers have the full responsibility for grading the students. There is no external evaluation of students, nor are there standardized tests that determine student grades. However, there are national tests in math, English and Swedish (or Swedish as a second language), with a common grading scheme, and the scores on these tests typically have a big weight in the final grade the student receives.

Importantly, the final upper-secondary school GPA is one of two existing screening tools for admittance to tertiary education, the other one being a college-admissions test. At least one third of the slots should be set aside for admittance based on GPA, and one third held in reserve for those who took the national university aptitude test. A high test score on the aptitude test is however not enough; the student is still required to have a minimum level in terms of grades in certain subjects. As a consequence, the course grades are important for the students.

IV. Empirical Analysis: The Effect of Same-Sex Teacher on Student Outcomes

A. The Data

The data used in this study stem from the database of upper-secondary education (HANNA), administered by the municipality of Stockholm. The database contains detailed information on students' grades and teacher and class assignment for 69 upper-secondary schools in the Stockholm area. For our purposes, the advantage of these data is that we can identify for each student and each course, both the student outcome (the final course grade), the gender of the teacher, and the class. We also attain information on teacher's age, school, class size, gender composition in the class, and the specific course programme from the database.

Our sample consists of students attending upper-secondary education in Stockholm municipality, graduating in 1997 – 2004. In particular, we restrict our sample to those who graduated within three years (the expected time), and who have no history of grade repetition. We further restrict our analysis to students taking any of the two main academic tracks, that is, the social science programme or the natural science programme.

Our motivation for this restriction is twofold. First, we want the sample to be homogenous with respect to student ability and motivation. This is mainly to make sure that teachers are not allocated to groups of students based on non-conformist behaviour of the students. We believe that this is unlikely to happen at the academic tracks of upper-secondary school, where students are motivated. Second, while we strive for homogeneity in student aspirations, we need heterogeneity in terms of the gender of both students and teachers. Many of the vocational tracks in the Swedish upper-secondary school are gender segregated (examples of such tracks are child caring and construction), meaning that we do not loose much in terms of identifying variation by dropping these from the sample.

We limit our analysis to compulsory courses within each study track, and we study only those subjects that require more than one course. This is because if courses are mandatory, students can not choose or opt out of the course based on teacher characteristics. We also need at least two courses by subject to be mandatory in order to explore within-student-subject variation. Thus, we study the relationship between student outcomes and same-sex teacher in the following subjects: mathematics, Swedish and English.

We use course grades as our measure of student performance. Ideally, we would have preferred to have both course grades and subject-specific test scores, to be able to separate discrimination and role model effects. Test scores are not available, but we argue that in the choice between the two, grades are actually a preferred outcome measure.[5] The reason is that the course grade includes also the teacher's subjective assessment of the student, something that we want to capture in our analysis.[6] We transform these grades into a numeric scale by assigning the values 0, 10, 15 and 20 to the different levels; these values correspond to the values that are used when summarizing all final course grades into a total GPA.

The independent variables in our analysis are gender dummies for the teacher and the student, an interaction between student's and teacher's gender, class size, share of

[5] Even though nation wide tests are given in the subjects of interest in this study, the test results are unfortunately not recorded, and hence not available.

[6] Note also, that if we are interested in studying discrimination, it is crucial to have an outcome of teacher assessment. Grades constitute such a measure, although they do not provide sufficient evidence for conclusions about discrimination.

female students in the class and age of the teacher. We also include school-specific effects and year effects, the latter in order to capture grade inflation, which seems to be present in Swedish upper-secondary school (Wikström and Wikström 2004). The independent variables and the empirical specification are discussed further in section IV.C.

The HANNA database is not primarily intended for research; the information is reported by teachers, principals and administrative personnel at the respective schools. Thus, many different persons collaborate in entering information into the system, and even if there are routines for this, there is obviously room for errors (for example, in some cases we miss information on teacher's age and gender). Most likely, these errors occur randomly but will nevertheless introduce measurement error.

Because of the misreporting in the data, we are forced to restrict our data further. The birth year indicator for students and teachers sometimes comes out as improbable, so we restrict the students to be born in 1978 – 1985, and the teachers to be between 22 and 70 years old. Moreover, in our data class size ranges from 1 to 40.[7] We believe that the very small class sizes are due either to reporting errors or to the fact that different subjects may be taught within the same class. To clean our data we allow class size to be no smaller than 10 and require that the subject should be the same throughout a specific class. We also drop multiple observations of the same course grade, keeping the first grade obtained.[8] Finally, we keep only the individuals who have completed the full sequence of courses within a subject. All in all, the above restrictions reduce the sample somewhat; almost 9 percent out of 17,744 unique individuals are dropped from our sample due to these restrictions.

B. A Descriptive Overview of the Gender Grade Gap

We start out by documenting the gender grade gap by course, for the courses in our sample. Table 1 reports average grades for each course, by student gender and by orientation of study. First, note that the grades are on average higher in Swedish and

[7] We construct a measure of class size by defining a class as those individuals taking the same course the same semester, in the same school with the same teacher and in the same course group.

[8] Multiple grades in the same course is the result of students retaking the course in order to improve their grade.

English than in math; this holds for both science and social science students, and for female and male students.[9] Next, and more importantly, the gender grade gap, defined as the difference between female and male students' grades (see columns 3 and 7), is the highest in Swedish and the lowest in math. The grade difference between female and male students in the social science programme is as high as 2.37 grade points for Swedish B, which corresponds roughly to half of a standard deviation in the distribution of that particular grade. Looking at the other extreme, we find the smallest gender grade differences in mathematics; most of them are still positive indicating a female advantage, but for natural science students, the first math course (math A) is associated with a negative gender gap favouring male students. Overall the picture is clear, however: in terms of course grades, girls dominate in all subjects and courses in our sample.

Our main interest in this study is to relate the gender grade gap to the gender composition of the teacher profession. To clarify this possible association, columns 4 and 8 in Table 1 report on the share of female teachers for the specific course. The share of female teachers is as high as 75 – 85 percent in Swedish, around 65 percent in English and around 35 – 45 percent in math. Thus, a striking pattern emerges when comparing columns 3 and 4, 7 and 8: the higher the share of female teachers, the larger is the gender grade gap. The remainder of our analysis focuses on exploring whether there is a causal mechanism explaining this relationship; and the causal mechanisms that we have in mind are those discussed in section I. Do same-sex teachers act as role models for their students, thereby enhancing performance of students with the same gender as the teacher? Or are there discriminatory practices in teacher's evaluation of the student, in the sense that teachers favour students of their own gender?

C. The Econometric Approach

The ideal research design that would allow us to study whether the gender school performance gap is related to teacher characteristics, is an experiment where students and teachers are randomly assigned to classes. This would ensure that there was no selection process at stake, and we would be confident that our findings were not plagued by selection bias. Further, we want to be able to distinguish between our two previously

[9] One exception being the high math A grade for the natural science programme.

described hypotheses; discrimination or role models? To be able to do this, test scores from a gender-blind and a non-gender-blind test, of the type described in Lavy (2004), are necessary. If there were a difference in the blind and the non-blind score, and if this difference were positively associated with same-sex teacher, we would conclude that when setting grades teachers are discriminating in favour of their own gender. The other alternative is that there is no difference between the blind and non-blind score, but still a positive association between same-sex teacher and outcomes. This is consistent with the role model hypothesis, but also with discrimination in teaching. If teachers discriminate in the time and effort they put into teaching, this will show up in the students' performance.

Our data set is different from the ideal design in the following way: since the data have not been generated by an experiment, we can not rely on random student-teacher matching, and we do not have access to a gender-blind test score. In addition, our measure of student performance, the final course grade, does not allow us to separate between the discrimination and the role model hypothesis. We are able to assess the extent to which a same-sex teacher enhances students' grades, but we cannot determine whether any such effect comes from the teacher's grading habits, from same-sex teacher induced changes in student performance, or from discrimination in teaching. Below, we discuss the implications of the deviation from the ideal research design; which assumptions we must rely on and how it affects the interpretation of our findings.

C.1 OLS

The OLS approach relies on the assumption that within a specific school and a specific subject, student-teacher assignment is random. There are several reasons to believe that this assumption is valid. First, we restrict our analysis to subjects and courses that are mandatory for students within a certain educational program. Therefore, choice of courses, based on e.g. teacher characteristics, should not be a confounder of our results. Second, since we are restricting our study to students in the academic tracks in upper-secondary level, the risk that the school assigns teachers to different groups of students based on non-conformist behaviour should be limited. The school has some information on student characteristics pre-enrolment, but only if the school chooses to sort students based on these characteristics and then assigns teachers based on this sorting, student-teacher assignment would be non-random. We have no evidence that sorting of this kind is taking place in Swedish upper-secondary schools. It is important to note that since we are assuming random student-teacher assignment within the subject-school cell, school choice (of both teachers and students) is not violating our results. Furthermore, gender differences in teacher quality should not be a confounder of our results; it is taken care of by our main effect of teacher's gender.[10] We estimate the following equation:

(1)

$$y_{isft} = \alpha + \beta_1 FEMSTUD_{isft} + \beta_2 FEMTEACH_{isft} + \beta_3 INTERACT_{isft} + \gamma' X_{isft} + \eta_s + \lambda_f + \mu_t + \varepsilon_{isft}$$

y_{isft} is the grade outcome for student *i*, in school *s*, in subject *f*, in year *t* (taking on the four different values 0, 10, 15 or 20). *FEMSTUD* is a dummy variable for female students, *FEMTEACH* is a dummy variable for the teacher being female, and *INTERACT* is an interaction term between female student and female teacher. *X* is a vector of controls containing the share of female students in the class, class size and teacher's age.[11] λ_f is a set of indicator variables for each subject, η_s and μ_t capture school-specific effects and year effects respectively.[12] ε_{isft} is the error term, that we assume is normally distributed and iid. The coefficient we are particularly interested in evaluating is the one on the interaction between female student and female teacher, β_3.

We estimate equation 1 first by pooling all our subjects and courses together. By doing so, we force the coefficients measuring effects on student outcome to be the same

---

[10] Gender differences in teacher quality will be a problem if for example high quality female teachers are matched with high performing female students and low performing male students, and high performing male teachers are matched with high performing male students and low performing female students. This type of matching seems to us unlikely.

[11] These control variables can be motivated in the following way: 1. The share of female students in the class captures peer effects. 2. Class size is a standard input variable in education production functions and is included to control for the fact that female and male teachers might be allocated to classes of different sizes. 3. Teacher's age is included to capture time trends in teacher quality.

[12] For efficiency reasons we pool courses within a subject; hence we pool the courses English A and English B to form the subject English. Likewise, we pool Swedish A and B, and math A-D for natural science students and math A and B for social science students.

across subjects. In particular, initially we force the pure student gender component (β_1) to be the same in all subjects, which is a priori unlikely.

Table 2 summarizes our main findings. (Descriptive statistics for the pooled samples are presented in Appendix A). Panel A presents results for natural science students, and panel B presents the corresponding results for social science students. Column 1 reports the main gender effects and the interaction between teacher and student gender, excluding all other control variables. The results are in line with the observed pattern in Table 1. All else equal, female students have higher course grades than males; this holds for both the natural and social science students. The effect, measured as the sum of β_1 and β_3, is markedly higher for the social science students. Moreover, in the social sciences, female teachers set higher grades than male teachers. The interaction term (female student*female teacher) is positive, indicating that having a teacher of the same sex increases student course grades by 0.3-0.65 grade points.

Column 1 in Table 2 allows a cross-subject and a cross-school comparison, not controlling for the fact that teacher sorting into different subjects and schools is non-random. In column 2 we now introduce control variables as specified in equation 1; most importantly we include school and subject-specific dummies. Thus, we assume that within a school and subject, student-teacher assignment s random. The results in column 2 show that *within subject*, the interaction term, β_3, indicates a positive effect of a same-sex teacher for social science students. The point estimate is 0.7 grade points, which corresponds to 13 percent of a standard deviation.

Can the results in column 2, Table 2, be taken as evidence of a causal effect of same-sex teacher on student outcomes? Teachers are not randomly assigned to subjects, in fact, as we concluded from Table 1, the share of female teachers is higher in subjects where girls are high performing. The specification in column 2 includes subject dummies, thus accounting for the non-random allocation of teachers to different subjects. The specification does not, however, take into account that the gender grade gap coefficient, β_1, may vary by subject. Since we do not let the main effect of student gender differ across subjects, any variation in the main gender effect across subjects that is correlated with teacher's gender, will introduce a spurious correlation between teacher's gender and student outcomes.

In column 3 we relax the assumption that the main gender effect is the same across subjects, by interacting female student and subject. This clearly has an effect on the parameter estimate of the student-teacher interaction term; the interaction term is no longer statistically different from zero. We conclude that the positive relationship between the gender gap and teacher's gender, that we observed in previous specifications, is purely an association caused by teacher sorting into different subjects. The logic to this is that we assume that teachers choose to teach in the subjects where they themselves are high performers, meaning that if there is some unobserved mechanism (we may think of it as culture) that make women perform better in humanities than in science, and this mechanism influenced the teacher generation in the same way as it influences today's students, teacher sorting will introduce a positive correlation between the performance and having a same-sex teacher.

Thus, at first glance (columns 1 and 2), it seems like within a certain subject, the female advantage is even larger when the teacher is female, which we could interpret as either an effect of role models or as discrimination. However, letting the female advantage vary by subject, the gender interaction term is statistically insignificant. We can not explain the female advantage in school performance by a positive effect of having a same-sex teacher.

C.2 Within-Student-Subject Estimation

Our previous empirical specification has a major limitation; it does not take into account that current achievement will be influenced by teacher and school characteristics in previous periods. If current and past teacher characteristics are correlated, we might expect the estimated effect of e.g., teacher's gender to be biased. To remedy this shortcoming, we explore the course structure of the Swedish upper-secondary school. During the three year program, students complete a sequence of courses within each subject, for example math A, math B, math C etc. By the end of each course, students are given a final course grade. Because of teacher turnover and student mobility, students might not be assigned the same teacher in all courses of the sequence, thus opening up for variation in teacher's gender, *within subject and student*. In essence, what we are doing is to examine whether grade gains over time can be attributed to a change from a teacher of the opposite sex to a same-sex teacher.

Apart from the issue of past teacher and school characteristics, the OLS estimates presented above will be biased also if the assumption of random student-teacher matching *within a subject* does not hold. This might be the case if, for example, female teachers are assigned groups with relatively high (or low) performing girls, and male teachers are assigned groups with relatively high (or low) performing boys. Even though we have no evidence of this type of sorting, our within-student-subject identification strategy takes care of any teacher sorting, since the student fixed effect controls for fixed student ability and motivation within the subject.

The within-student-subject variation may well be small, but it allows us to control both for unobserved ability of the student, and importantly, for relevant influences of previous periods.[13]

We estimate the following equation:

(2)

$$y_{isft} = \alpha + \beta_1 FEMTEACH_{isft} + \beta_2 INTERACT_{isft} + \gamma' X_{isft} + \eta_s + \lambda_f + \omega_i + \mu_t + \varepsilon_{isft}$$

Note that equation 2 is similar to equation 1, but importantly, individual fixed effects ω_i are introduced. School and year effects are still included to control for student mobility and timing of the course.

As a first step of the within-student-subject analysis we turn to the question of identifying variation in the data. We rely on the gender of the teacher changing over time, either because of teacher turnover, student mobility or the course allocation among teachers within a school. In Appendix B (Table B1) we explore the structure of our identifying variation. First, we see that the transition probabilities for a change in teacher's gender range from 13 to 67 percent of the respective sub-samples of our data. Thus, there is indeed some within-student-subject variation in the sex of the teacher that

[13] This specification is similar to a value-added specification (see for example Hanushek 1986 and Todd and Wolpin 2003). Such a specification regresses achievement gain between two periods on current explanatory variables, thereby netting out the influences of previous periods on current achievement. We use the panel structure of our data and let the change in teacher's gender identify the coefficient, as opposed to the current teacher characteristic.

we explore below. Second, for an unbiased estimate, the change of teacher must be random with respect to the sex of the student. That is, we want to rule out that the student-teacher interaction is correlated with previous teacher characteristics. Table B1 presents the transition probabilities for female and male students pooled together, but importantly, those probabilities do not differ much by student gender.[14] We take this as evidence that there is no systematic difference in change of teacher for female and male students. Moreover, the reason that a student experiences different teachers in different courses can be either that the student moves (change of class, change of school), teacher turnover (change of class, change of school, retirement) or that the teacher does not teach all courses. In our sample, most changes of teacher occur for the whole class, meaning that our identifying variation mainly comes from teacher turnover.[15] This is comforting since it means that within a class there is no selection among the students with respect to change of teacher.

Columns 4 and 5 in Table 2 present the within-student-subject estimates. In column 4, we hold the main effect of student gender constant across subjects. Just as in the corresponding OLS estimates (column 2), there is a statistically significant positive effect of having a same-sex teacher, in particular for social science students. Moving from column 4 to column 5, where we allow the main gender effect to vary across subjects, we see that the observed effect in column 4 is driven by a spurious correlation between teacher's gender and the main gender grade gap. The last column strengthens us in concluding that there is no effect on grades of having a same-sex teacher in Swedish upper-secondary school.

C.3 Subject-specific estimates

As an extension to our previous results, we present also estimates by subject. Our pooled regressions might mask that the effects are different across subjects. Table 3 shows cross-sectional and within-student estimates for English, Swedish and mathematics, respectively. We estimate both positive and negative interaction effects, most of which

[14] The transition matrices by gender are not presented in the paper but can be obtained from the authors upon request.

[15] For natural science students, of all teacher changes, 95, 95 and 98 percent are due to teacher turnover for English, Swedish and math, respectively. The corresponding numbers for social science students are 95, 93 and 86 percent. (Teacher turnover is here defined as when all students in the class change teacher).

with large standard errors. In one case we find a statistically significant effect of having a same-sex teacher; that is, for natural science students, a same-sex teacher improves their Swedish grade by 0.6 grade points. 0.6 grade points corresponds to a tenth of a standard deviation in the Swedish grade variable, and we should bear in mind that this is the effect of changing to a same-sex teacher in one course, one subject. The effect of an increase of 0.6 grade points in a Swedish course will therefore have a relatively mild impact on the overall GPA of the student.

V. Conclusions

It is an international phenomenon that girls perform better than boys in school. This is particularly true in some subjects, for example languages, but also in previously male dominated subjects, like science and mathematics, girls are now doing better and their scores are moving closer to those of their male peers. In this paper, we investigate whether we can explain the gender grade gap in school performance with the sex of the teacher, that is, is it beneficial for the students to be taught by a teacher with the same sex? We lay out two possible hypotheses as to why the gender of the teacher may influence student outcomes. First, one explanation may be that teachers favour students that are more like themselves, whereby they rank students of their own gender higher on the margin. Second, teachers may constitute role models for the students, thus, having a same-sex teacher may affect the student's effort and therefore her or his outcome.

Using data from Swedish upper-secondary school, we estimate both OLS and fixed effects regressions. In particular, we explore the course structure of Swedish upper-secondary school, where students take a sequence of courses in the same subject. Each course in a sequence is graded separately, and our source of identification stems from the fact there is a significant amount of teacher turnover and student mobility across courses. This enables us to control for individual ability effects, teacher sorting into different subjects, and the influence of school inputs in previous periods on the current outcome.

Our results show a clear association between the female grade advantage and female teachers, but do not support our hypothesis that a same-sex teacher has a positive causal impact on student outcomes, measured in terms of course grades in upper-secondary school. Importantly, the results laid out in this paper are not in line with those in Dee (2005a, 2005b); Dee finds that a same-sex teacher indeed has an impact on student performance. Future research will hopefully return to this topic to further make us understand the importance of matched student-teacher characteristics.

Our results also shed light on an important policy issue; to what extent should policy makers try to influence the gender composition of the teacher profession? In the case of strong results pointing at role models as being important, or if discrimination is prevalent, there is room for an education policy that aims at gender balance in recruitment of new teachers. In this study, however, we find no evidence that can motivate such a policy. On the contrary, it seems like teacher's gender have no effect on student outcomes.

There is also a broader lesson to be learned from our findings, a lesson related to methodology. We can think of several other applications with similar sources of bias and where the methodological issues are comparable. For example, consider the literature that relates the gender wage gap in an establishment to the sex of the manager/boss (Hultin and Szulkin 2003). Hultin and Szulkin find that the gender wage gap is increasing in the share of male managers and supervisors in an establishment. However, it is likely that female and male managers select into somewhat different types of establishments, whereby we cannot conclude that the estimated relationship is causal. This is similar to our argument that female and male teachers sort into different subjects, meaning that we cannot take our initial OLS results as evidence of a causal effect of having a same-sex teacher. When we extend our analysis to take into account this type of sorting, we find no evidence of a causal effect. That is, in other similar applications it is important to consider that once this selection issue is taken care of, the findings might come out as less robust.

We close by laying out a few possible explanations to why we find that the teacher's gender has no effect. First, Dee (2004, 2005a, 2005b) studies the effects of same-race and same-sex teachers for younger children (kindergarten through 3rd grade and 8th grade), whereas we focus on 16-18 year-olds. It may be the case that gender effects are prevalent, but that they are more important at an early stage in the child's education, which might explain why we do not detect any effect that we can label as causal at upper-secondary level. Second, we are looking at a group of relatively motivated students. If we think that role models play a role in the education production function, their effect might not be linear across the distribution of student abilities or student's family background.[16] If role models are more important for less able or less motivated students, we fail to capture this by studying theoretical programmes at the upper-secondary level. Third, our grade variable might not be informative enough to capture variations in student performance or teacher preferences. The grade scale contains only four steps; possibly a finer measure would improve our analysis.

[16] Dee (2005a) finds that the interaction effect between student and teacher demographics is much stronger for children with low socio-economic status than for children from more advantaged backgrounds.

VI. References

Ashworth, John and J. Lynne Evans (2001), “Modeling Student Subject Choice at Secondary and Tertiary Level: A Cross-Section Study”, Journal of Economic Education 32(4): 311-320.

Bettinger, Eric P. and Bridget Terry Long (2005), “Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students”, *The American Economic Review,* forthcoming.

Campbell, Jay R., Catherine M. Hombo and John Mazzeo (1999), “Trends in Academic progress: Three decades of student performance”, National Center for Education Statistics 2000-469.

Canes, Brandice J. and Harvey S. Rosen (1995), “Following In Her Footsteps? Faculty Gender Composition and Women’s Choices of College Majors”, *Industrial and Labor Relations Review* 48(3): 486-504.

Dee, Thomas S. (2004), “Teachers, Race and Student Achievement in a Randomized Experiment”, *The Review of Economics and Statistics* 86(1): 195-210.

Dee, Thomas S. (2005a), “A Teacher Like Me: Does Race, Ethnicity or Gender Matter?”, *American Economic Review*, 95(2).

Dee, Thomas S. (2005b), “Teachers and the Gender Gaps in Student Achievement”, NBER Working Paper 11660.

Dynan, Karen E. and Cecilia Elena Rouse (1997), “The Underrepresentation of Women in Economics: A Study of Undergraduate Economics Students”, *Journal of Economic Education* 28(4): 350-368.

Ehrenberg, Ronald G. and Dominic J. Brewer (1995), "Did Teachers' Verbal Ability and Race Matter in the 1960s? *Coleman* Revisited", *Economics of Education Review* 14(1): 1-21.

Ehrenberg, Ronald G., Daniel D. Goldhaber and Dominic J. Brewer (1995), "Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988", *Industrial and Labor Relations Review* 48(3): 547-561.

Hanushek, Eric A. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature* 14: 1141-1177.

Hanushek, Eric A., John F. Kain and Steven G. Rivkin (2004), "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievment", Working Paper.

Hultin, Mia and Ryszard Szulkin (2003), "Mechanisms of Inequality: Unequal Access to Organizational Power and the Gender Wage Gap", *European Sociological Review* 19(2): 143-159.

Lavy, Victor (2004), "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence From a Natural Experiment", NBER Working Paper 10678.

Neumark, David and Rosella Gardecki (1988), "Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics", *The Journal of Human Resources* 33(1): 220-246.

Rask, Kevin N. and Elisabeth M. Bailey (2002), "Are Faculty Role Models? Evidence from Major Choice in an Undergraduate Institution", *Journal of Economic Education* 33(2): 99-124.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain (2005), “Teachers, Schools and Academic Achievement”, *Econometrica*, forthcoming.

Rothstein, Donna S (1995), “Do Female Faculty Influence Female Students’ Educational and Labor Market Attainments?”, *Industrial and Labor Relations Review* 48(3): 515-530.

Solnick, Sara J. (1995), “Changes in Women’s Majors from Entrance to Graduation at Women’s and Coeducational Colleges”, *Industrial and Labor Relations Review* 48(3): 505-514.

Statistics Sweden (SCB) (1994), ”Utbildningsstatistisk årsbok 1995”, Statistiska Centralbyrån.

Statistics Sweden (SCB) (2001), ”Utbildningsstatistisk årsbok 2002”, Statistiska Centralbyrån.

Statistics Sweden (SCB) (2004), ”Utbildningsstatistisk årsbok 2005”, Statistiska Centralbyrån.

Steele, Claude M (1997), “A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance”, *American Psychologist* 52(6): 613-629.

Swedish National Agency for Education (Skolverket) (2004), “Utbildningsresultat 2004, Riksnivå – Sveriges officiella statistik om förskoleverksamhet, skolbarnomsorg, skola och vuxenutbildning”, part 1.

Todd, Petra E. and Kenneth I. Wolpin (2003), “On the Specification and Estimation of the Production Function for Cognitive Achievement”, *The Economic Journal* 113: F3-F33.

Wikström Christina and Magnus Wikström (2004), “Grade inflation and school competition: an empirical analysis based on the Swedish upper-secondary schools”, *Economics of Education Review* 24: 309-322.

Table 1

The gender grade gap by course

Course	Natural science programme: Average grade (Std. dev)		Difference (Std. err)	% female teachers	Social science programme: Average grade (Std. dev)		Difference (Std. err)	% female teachers
	(1) Female student	(2) Male student	(3)	(4)	(5) Female student	(6) Male student	(7)	(8)
English A	17.05	16.64	0.41***	0.63	15.58	14.69	0.90***	0.66
	(3.21)	(3.68)	(0.1)		(4.09)	(4.18)	(0.11)	
English B	15.35	14.86	0.48***	0.61	13.57	12.59	0.98***	0.64
	(4.1)	(4.6)	(0.13)		(4.93)	(5.25)	0.13	
Swedish A	16.58	14.9	1.69***	0.75	14.84	12.92	1.92***	0.85
	(3.42)	(3.96)	(0.11)		(4.12)	(4.36)	(0.11)	
Swedish B	16.65	14.9	1.75***	0.75	15.01	12.63	2.37***	0.80
	(3.89)	(4.43)	(0.12)		(4.58)	(5.24)	(0.13)	
Math A	16.25	16.49	- 0.24**	0.45	12.58	12.39	0.20**	0.42
	(3.8)	(3.48)	(0.09)		(4.57)	(4.23)	(0.1)	
Math B	14.84	14.28	0.56***	0.47	9.86	8.95	0.91***	0.34
	(4.80)	(5.09)	(0.13)		(6.19)	(6.02)	(0.13)	
Math C	14.07	13.81	0.26**	0.45				
	(5.14)	(5.22)	(0.13)					
Math D	12.64	12.34	0.30**	0.43				
	(5.88)	(6.02)	(0.15)					

*Note: ***/** denote significance at the 1/5 percent level, respectively.*

Table 2

Pooled OLS and within-student estimates

Dependent variable: course grade

Explanatory variable	(1) OLS	(2) OLS	(3) OLS	(4) Within-student	(5) Within-student
A. Natural science programme					
Female student	0.447	0.020	-0.223		
	(0.129)***	(0.109)	(0.110)**		
Female teacher	0.134	-0.048	0.064	-0.160	-0.026
	(0.234)	(0.176)	(0.164)	(0.161)	(0.145)
Female student*Female teacher	0.306	0.248	-0.044	0.310	-0.041
	(0.180)*	(0.168)	(0.137)	(0.159)*	(0.122)
Other controls	No	Yes	Yes	Yes	Yes
Subject*Female student	No	No	Yes	No	Yes
N	43,916	43,916	43,916	43,916	43,916
R^2	0.00	0.08	0.08	0.55	0.55
B. Social science programme					
Female student	0.753	0.294	0.029		
	(0.157)***	(0.105)***	(0.121)		
Female teacher	0.902	-0.241	0.029	-0.252	0.091
	(0.316)***	(0.207)	(0.215)	(0.185)	(0.193)
Female student*Female teacher	0.653	0.703	0.251	0.591	0.015
	(0.210)***	(0.158)***	(0.163)	(0.146)***	(0.146)
Other controls	No	Yes	Yes	Yes	Yes
Subject*Female student	No	No	Yes	No	Yes
N	42,624	42,624	42,624	42,624	42,624
R^2	0.03	0.16	0.17	0.60	0.61

*Notes: Standard errors, in parentheses, are clustered on teachers. Other controls refers to controls for class size, share of females in the class, teacher's age, and dummies for subject, school and year. */**/*** refer to statistical significance at the 10, 5 and 1 percent level respectively.*

Table 3

OLS and within-student estimates

Dependent variable: course grade

Explanatory variable	English (1) OLS	English (2) Within-student	Swedish (3) OLS	Swedish (4) Within-student	Math (5) OLS	Math (6) Within-student
A. Natural science programme						
Female student	-0.101		1.270		-0.242	
	(0.181)		(0.206)***		(0.128)*	
Female teacher	-0.147	-0.052	-0.473	-0.499	0.190	0.081
	(0.251)	(0.302)	(0.203)**	(0.224)**	(0.212)	(0.190)
Female student*Female teacher	0.198	-0.055	0.307	0.600	-0.176	-0.143
	(0.241)	(0.340)	(0.247)	(0.227)***	(0.185)	(0.185)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
N	9,578	9,578	9,578	9,578	24,760	24,760
R^2	0.10	0.85	0.12	0.80	0.06	0.70
B. Social science programme						
Female student	0.155		1.368		0.139	
	(0.187)		(0.295)***		(0.135)	
Female teacher	-0.224	0.032	-0.620	0.090	0.398	0.171
	(0.282)	(0.339)	(0.344)*	(0.328)	(0.347)	(0.293)
Female student*Female teacher	0.196	0.294	0.508	0.043	0.246	-0.226
	(0.215)	(0.268)	(0.323)	(0.309)	(0.265)	(0.220)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
N	12,734	12,734	12,734	12,734	17,156	17,156
R^2	0.14	0.85	0.13	0.81	0.10	0.78

Notes: Standard errors, in parentheses, are clustered on teachers. Other controls refers to controls for class size, share of females in the class, teacher's age, and dummies for school and year./**/*** refer to statistical significance at the 10, 5 and 1 percent level respectively.*

Appendix A

Table A1

Descriptive statistics
Pooled samples

Variable	(1) Natural science programme Mean (Standard deviation)	(2) Social science programme Mean (Standard deviation)
Course grade	14.96	12.87
	(4.80)	(5.30)
Female student	0.42	0.59
	(0.49)	(0.49)
Female teacher	0.55	0.59
	(0.50)	(0.49)
Interaction	0.23	0.35
(Female student*Female teacher)	(0.42)	(0.48)
Class size	27.57	28.10
	(3.90)	(4.35)
Share female students in class	0.42	0.58
	(0.19)	(0.16)
Teacher's age	51.06	48.89
	(10.80)	(10.98)
Swedish A indicator	0.11	0.15
	(0.31)	(0.36)
Swedish B indicator	0.11	0.15
	(0.31)	(0.36)
English A indicator	0.11	0.15
	(0.31)	(0.36)
English B indicator	0.11	0.15
	(0.31)	(0.36)
Math A indicator	0.14	0.20
	(0.35)	(0.40)
Math B indicator	0.14	0.20
	(0.35)	(0.40)
Math C indicator	0.14	
	(0.35)	
Math D indicator	0.14	
	(0.35)	
Year	1999.37	1999.70
	(2.09)	(2.08)
n	43,916	42,624

Appendix B

Table B1

Transition matrices

A. Natural science students		*Teacher's gender in second course*		
English				
		Woman	Man	Total
Teacher's gender in first course	Woman	2,476	557	3,033
		81.62	18.36	100.00
	Man	449	1,307	1,756
		25.57	74.43	100.00
	Total	2,925	1,864	4,789
		61.08	38.92	100.00
Swedish				
Teacher's gender in first course	Woman	2,767	804	3,571
		77.49	22.51	100.00
	Man	813	405	1,218
		66.75	33.25	100.00
	Total	3,58	1,209	4,789
		74.75	25.25	100.00
		Teacher's gender in period t+1 course		
Math				
Teacher's gender in period t course	Woman	6,098	2,339	8,437
		72.28	27.72	100.00
	Man	2,203	7,930	10,133
		21.74	78.26	100.00
	Total	8,301	10,269	18,570
		44.70	55.30	100.00

Table B1, cont.

Transition matrices

B. Social science students		*Teacher's gender in second course*		
English				
		Woman	Man	Total
Teacher's gender	Woman	3,623	556	4,189
in first course		86.49	13.51	100.00
	Man	446	1,732	2,178
		20.48	79.52	100.00
	Total	4,069	2,298	6,367
		63.91	36.09	100.00
Swedish				
Teacher's gender	Woman	4,503	925	5,428
in first course		82.96	17.04	100.00
	Man	576	363	939
		61.34	38.66	100.00
	Total	5,079	1,288	6,367
		79.77	20.23	100.00
Math				
Teacher's gender	Woman	1,874	1,694	3,568
in first course		52.52	47.48	100.00
	Man	1060	3950	5010
		21.16	78.84	100.00
	Total	2,934	5,644	8,578
		34.20	65.80	100.00