Neugebauer, Martin; Helbig, Marcel; Landmann, Andreas

Unmasking the Myth of the Same-Sex Teacher Advantage

formal überarbeitete Version der Originalveröffentlichung in:
formally revised edition of the original source in:
European sociological review 27 (2011) 5, S. 669-689

urn:nbn:de:0111-opus-79552

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in European Sociological Review following peer review. The definitive publisher-authenticated version Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the Myth of the Same-Sex Teacher Advantage. European Sociological Review, 27(5), 669–689 *is available online at: http://esr.oxfordjournals.org/content/27/5/669.full.pdf+html*

DOI:10.1093/esr/jcq038, Online publication 20 July 2010

Unmasking the Myth of the Same-Sex Teacher Advantage

Martin Neugebauer, Marcel Helbig and Andreas Landmann

Abstract. *Trend statistics reveal a striking reversal of a gender gap that has once favoured males: Girls have surpassed boys in many aspects of the educational system. At the same time, the share of female teachers has grown in almost all countries of the western world. There is an on-going, contentious debate on whether the gender of the teacher can account, in part, for the growing educational disadvantage of males. In this study, we use large-scale data from IGLU-E, an expansion of PIRLS in Germany, to estimate whether there is a causal effect of having a same-sex teacher on student outcomes. We estimate effects for typical 'female' subjects and typical 'male' subjects as well as for different student outcomes (objective test scores and more subjective teacher's grades). We find virtually no evidence of a benefit from having a same-sex teacher, neither for boys nor for girls. These findings suggest that the popular call for more male teachers in primary school is not the key to tackle the growing disadvantage of boys.*

Keywords: Education, Gender Inequality, Teacher Gender, Boy Crisis, Feminization of School

Authors' Addresses

Martin Neugebauer (to whom correspondence should be addressed)

Mannheim Center for European Social Research (MZES), University of Mannheim, 68131 Mannheim, Germany. Email: martin.neugebauer@mzes.uni-mannheim.de

Marcel Helbig

Social Science Research Center Berlin (WZB), Reichpietschufer 50, 10785 Berlin, Germany. Email: marcel.helbig@wzb.eu

Andreas Landmann

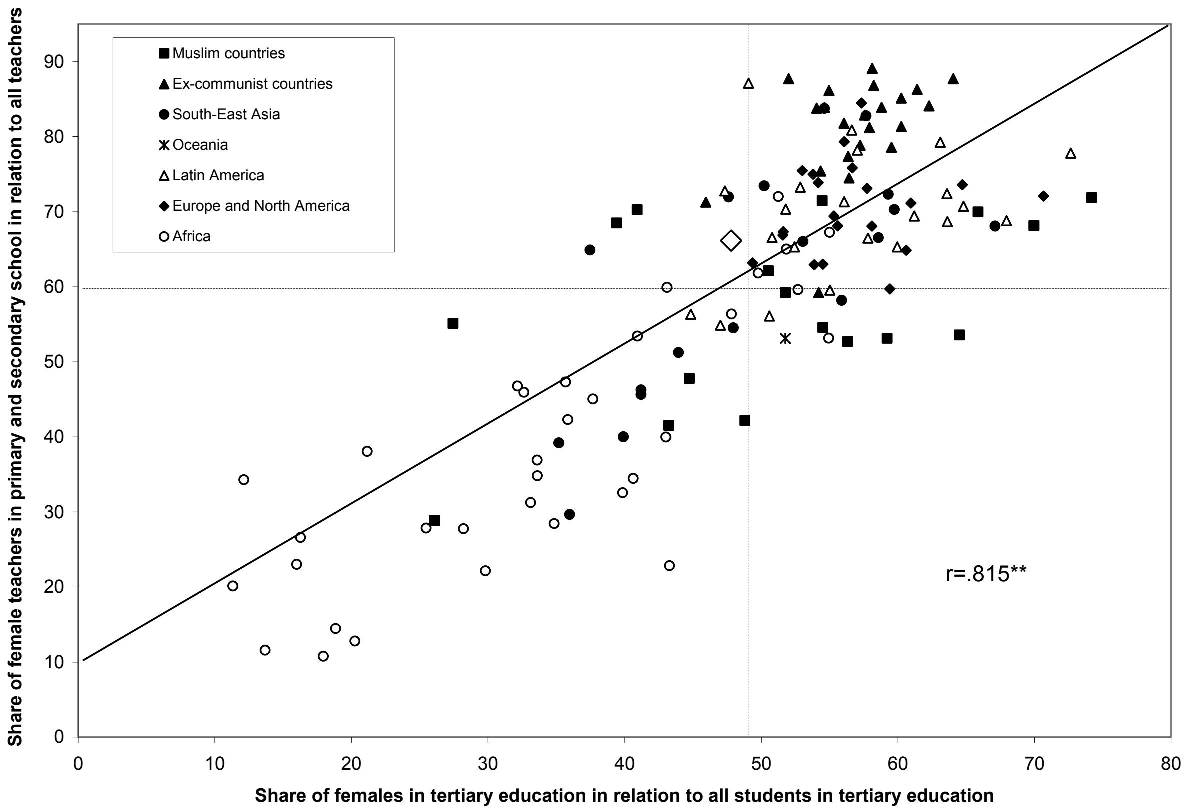
Center for Doctoral Studies in Economics (CDSE), University of Mannheim, 68131 Mannheim, Germany. Email: andreas.landmann@uni-mannheim.de

Introduction

In past decades, school systems in nearly all countries of the world have seen a gender-specific change in educational achievement. Even though girls' performance in math and science continues to lag behind that of boys in most countries, girls have been catching up with boys regarding both their math and science skills (Baker and Jones, 1993; Cole, 1997; Hedges and Nowell, 1995; IES, 2009; Willingham and Cole, 1997). What is more, girls have higher reading skills than boys and significantly higher writing skills (Cole, 1997; German PISA Consortium, 2007; Hedges and Nowell, 1995; IES, 2009; Wagemaker et al., 1996; Willingham and Cole, 1997). In nearly all countries of the world, women's participation in tertiary education, when compared to that of men, is at much higher levels than it was 40 years ago (UNESCO, 2009). In some European countries, the share of women among first-year university students has increased to more than 60 percent of all first-year students (OECD, 2009: own calculations). This development has led some researchers to talk about a "boy crisis" (Dammasch, 2007; Pollack, 2006) or even a "war against boys" (Sommers, 2000). As a consequence, girls' and women's most recent levels of educational success in relation to boys and men are interpreted as male educational failure. Research on this rare case of an inversion of a pattern of social stratification (Buchmann and DiPrete, 2006; Quenzel and Hurrelmann, 2010) is still in its fledgling stages, however. This is because a change of perspective with regard to gender-specific educational achievement only occurred in the late 1990s, and the causes of this new phenomenon have only been studied in recent years.

At the same time that girls' opportunities for educational success have been on the increase compared to those of boys, the share of women among the teaching population has been rising in virtually all countries of the Western hemisphere (Eurostat, 2009). This so-called "feminization of schooling" has been repeatedly linked to boys' poor educational opportunities when compared to girls (BBC-News, 2000; Dee, 2007; Sexton, 1969). It is conjectured that girls and boys tend to perform better with teachers of their own gender and/or receive better grades than with teachers of the opposite gender. It is assumed that girls have benefited academically from the increased share of female teachers among the teaching population, whereas boys have been negatively affected by the simultaneous decrease in male teachers at school. At the aggregate level, this hypothesis seems to be plausible. Figure 1 shows the correlation between the share of female teachers in primary and secondary schools in relation to all teachers and the share of female students enrolled in tertiary education in relation to all students in tertiary education in 135 countries. The two indicators are highly correlated - The higher the share of female teachers in a given country, the higher the share of female students who are enrolled in tertiary education.[1]

***Figure 1**: Correlation between the share of female teachers in primary and secondary schools in relation to all teachers and the share of females into tertiary education in relation to all students in tertiary education in 135 countries (2007).*



Source: (own figure based on UNESCO, 2010). The figure includes all countries for which data was available. Countries with a current tertiary student population of fewer than 3,000 students were not included. The horizontal and vertical dashed lines indicate the respective median value for all countries. The diagonal line represents the regression line. The bigger white square close to the median values represents Germany.

So far however, robust empirical evidence to confirm – at the individual level – the strong association which may be observed at the country level has been scarce. This lack of empirical support notwithstanding, efforts have been made to increase the share of male teachers, most importantly in Anglo-American countries (Driessen, 2007). This is considered an effective instrument for improving the performance of boys. In the United Kingdom, the Training and Development Agency for Schools started an initiative to encourage more men to become teachers (TDA, 2009), as did the Federal Government of Australia (ABC-News, 2004).

In this paper, we seek to examine whether teacher gender in fact has an impact on the academic achievement of male and female students, using the German supplemental sample from the 2001 PIRLS study (Progress in International Reading Literacy).[2] The evidence presented here aims to make several contributions to existing research. First, this study focuses on Germany, for which teacher gender effects have not yet been estimated directly (i.e. with individual-level data). Second, in addition to "gender-blind" performance scores, the educational outcomes in this study include teacher grades. Comparing grades and performance outcomes indicates whether teachers' gender in fact influences students' actual performance development or rather relates to their assessment of students, net of performance.[3] Third, we estimate effects across several subjects (German, Mathematics, and Science). Concentrating on these three domains is a reasonable approach because reading may generally be called a "girls' domain", whereas math is considered to be more of a "boys' domain".[4] Fourth, the rich data set allows us to estimate effects not only for the student population as a whole, but also for various subgroups, such as students from migrant or low socio-economic status families. Fifth, and perhaps most importantly, this study adopts an identification strategy that exploits the distinct features of the German primary education system, thus allowing us to (arguably) come very close to the "true" teacher gender effect: In Germany, primary school students typically stay with the same teacher for two or more years. More than 50 percent of primary school students in our sample even stayed with the same teacher for four years (that is, their entire school career up to this point). In other words, the majority of students in our data set were taught by only one teacher per subject. In addition, many students at primary school are taught by the same teacher in all major subjects. As a result, we may attribute students' learning process to that one teacher. To our knowledge, all previous studies encountered the problem that information on the gender or other characteristics of previous teachers was usually unavailable. In fact, in most instances, it was not even clear how long the teacher under examination had been teaching a student. This problem may partially account for the low degree of effects of teachers' gender on students' gender reported in previous studies. A second important feature of the German primary school system is the fact that school choice, for both teachers and students, is very limited. At primary school level, children are assigned to a school within their school district and may not freely choose an institution. Private schools are practically non-existent at this level. Furthermore, the procedure through which teachers apply to schools used to be very indirect. Up until the mid-1990s, new teachers were usually appointed by a central, state-level agency and then assigned to a school anywhere within the federal state in which they were applying (Treptow and Rothland, 2005).[5] As a result, the chance of being taught by a male versus female teacher should not depend on either the student's or the school's characteristics. Thus, student-teacher-gender pairing comes very close to a natural experiment.

Theoretical Considerations and Empirical Evidence

The feminization of schooling discourse is by no means a new phenomenon in Western societies. In the United States, where the teaching profession became a predominantly female profession as early as in the late 19th century, initial efforts to recruit more men into teaching were made between the two world wars (Blount, 2000; Martino, 2008; Rury, 1989). The main argument for doing so was the call for male role models, driven by an anxiety about the "feminization of boys", who would be less able to develop their masculinity without appropriate role models. This line of argument also characterized discussions in the 1960s (Sexton, 1969). In past as well as more recent contributions on the "feminization of schooling," the main focus of study is on boys. It is predominantly boys, according to this point of view, whose development is negatively affected by the absence of male teachers and whose academic performance, as a result, is weaker than that of girls. This interpretation tends to forget, however, that the relative loss of boys may also be explained by the rising educational success of girls. Gender-specific changes in educational success, therefore, may just as well be the result of a positive impact of female teachers on girls. This possibility is often neglected, however.

Today, there are several lines of argument that consider the feminization of schooling to have a negative impact on the educational performance of boys. The first line of argument points to a *lack of male role models* in school (Budde, 2008; Driessen, 2007; Holmund and Sund, 2008). From this point of view, the main complaint concerns the increasing absence of men in all stages of the education process, which results in a situation in which boys are unsure about their identity, lack clear patterns of gender role orientation (Budde, 2008: 492), and, as a result, are unable to form positive ideas of masculinity (Bacher *et al.*, 2008). This is closely related to the “same-sex hypothesis”, which is based on the idea that girls imitate women and that boys need men in order to eventually become men themselves (Powell and Downey, 1997). Psychology-based learning theory assumes that children tend to observe and identify with same-sex agents of socialization (Hannover, 2008; Kohlberg, 1966). At the same time, gender identification needs to be viewed in a bipolar way. Teachers, according to this theory, are better able to identify with a child of their own gender, possibly because they feel more competent in responding to their problems, having experienced them themselves (Powell and Downey, 1997). However, the link between gender identity and educational performance is inherently inconclusive (Blossfeld *et al.*, 2009). And yet, scholars, educators, and policy-makers keep using it for explaining gender-specific educational achievement.

Another possibility for explaining the gender achievement gap is a phenomenon referred to as *stereotype threat* (cf. Steele, 1997). According to this theory, students tend to perform less well when they fear they are being evaluated through the lens of a negative stereotype about the group they belong to. Teacher gender may enforce such negative stereotyping. If the math teacher is male, for example, girls may underperform simply because they believe math is a male domain. The negative stereotype is deactivated, however, when the teacher herself is female. Conversely, the same negative stereotype may be activated when boys take reading and writing classes with a female teacher.

Furthermore, there is some evidence that the *context in which a task is presented*, which may refer to the world of boys or the world of girls, leads to differences in the performance of boys and girls (Walther *et al.*, 2008). Math problems, for example, may be presented in a female context (using girls' names, pets, or cakes) or in a male context (using boys' names, buildings, trains, or income). If female teachers generally tended to choose female contexts and male teachers tended to choose male contexts to frame curricular content, they would thereby favor students of their own gender.

If all or any of the aforementioned arguments holds true, (H1:) boys' academic skills should be higher when they were taught by a male teacher, and girls' academic skills should be higher when they were taught by a female teacher.

Empirically, Dee (2007) was able to show that boys and girls have higher skills when they were taught by a same-sex teacher. With school-level aggregate data, Helbig (2010) did find positive effects regarding the reading skills of girls at schools with a higher share of female teachers, but he could not confirm the same-sex hypothesis with regard to boys' reading skills and the math skills of both genders. Ammermüller and Dolton's (2006) findings for England suggest that boys at the age of 13 do benefit somewhat from male math teachers, whereas girls benefit from female English teachers. They did not find significant effects in the US, however. In addition, they did not find any positive same-sex teacher effects for students around the age of 9. Likewise, several other authors were unable to come up with convincing empirical evidence for the claim that students benefit from being taught by a same-sex teacher with regard to their cognitive skills (Driessen, 2007; Ehrenberger *et al.*, 1995; Sokal *et al.*, 2007).

Another line of argument relates to the gender-specific evaluation of students, net of their actual performance. It is argued that female teachers may expect and reward certain types of behavior that girls have acquired in the process of socialization, whereas boys have not. While there is the possibility that female teachers actively discriminate against boys, the disadvantages that boys may experience from being taught by female teachers are more likely to be unintended consequences of the actions of female teachers, who may *interpret and judge boys' and girls' behavior in different ways* (Diefenbach and Klein, 2002). Types of behaviors that tend to disrupt school activities and negatively affect students' academic performance are more frequently found to originate from boys than from girls (Eagly and Chrvala, 1986). Possibly, female teachers are more "upset" with such types of behaviors than male teachers because their standards of judgment are derived from their own gender-specific socialization. In other words, one might say there is a certain mismatch between the "habitus" of female teachers and the "habitus" of male students (Brandes, 2002). According to this interpretation, feminization puts boys at a disadvantage especially when it comes to assessing their performance (Stamm, 2008; Holmund and Sund, 2008).

Gender influences on grades (net of performance) may also be caused by students behavior instead of teacher evaluation. According to the same-sex hypothesis, students taught by same-sex teachers may be more 'well-behaved' (e.g. participation in school lessons, disruption during class, doing homework) than students taught by other-sex teachers. As a consequence, the fact that students tend to get better grades from same-sex teachers would result more from a better behavior on the part of the student than from a conscious or subconscious grading bias in the teacher.

Based on these theoretical implications, (H2:) the grades boys earn from male teachers should be better than those they earn from female teachers, whereas girls should get better grades from female teachers than from male teachers. Previous research offers mixed results in this regard, too. With respect to grading, Ehrenberger at al. (1995), Hopf and Hatzichristou (1999), and Helbig (2010) were able to show that boys receive inferior grades from female teachers. The correlations that were identified must be considered quite weak, however.[6] What is more, a same-sex teacher advantage could not be confirmed by other studies, neither with regard to boys and girls (Carter, 1952; Driessen, 2007; Holmund and Sund, 2008) nor with regard to girls (Helbig, 2010).

In the following, we will assess whether girls or boys benefit from being taught by a same-sex teacher with respect to their academic performance (H1) and with respect to their grades, net of performance (H2).

Data and Variables

35 countries participated in the Progress in International Reading Literacy Study (PIRLS). This study was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 2001 to assess the reading literacy of nine- and ten-year-old students (Mullis *et al.*, 2003; Mullis *et al.*, 2002). Germany embedded this study within its national IGLU-E study, which focused on the reading and writing ability of fourth graders, as well as on their performance in math and science.[7] The PIRLS background questionnaires that were administered to students, parents, teachers, and schools were also expanded in IGLU-E. These modifications required a second test day to gather the information without overburdening the students (Schwippert, 2007). 12 out of 16 states in the Federal Republic of Germany participated in the extended PIRLS (IGLU-E). Data was collected in a two-stage stratified sampling design. First, participating schools were chosen. Within each school, a sample of classes from the targeted grade (fourth grade) was drawn. Within each class, in principle, all students were sampled. In practice, however, the number of sampled students was sometimes smaller than the actual class size because of student non-participation. After applying weights to compensate for non-proportionality and to ensure representativeness, the sample consisted of 5,858 students from 166 schools (308 classes).[8] The participation rate was satisfactorily high with 84% (details: Bos *et al.*, 2004; Bos *et al.*, 2003).[9] In the data set, extensive information on student performance, as well as on home and school environment is available through student, parent, teacher, and school questionnaires.

Dependent Variables. To test whether boys and girls actually know more when being taught (for at least 2 years) by a same-sex teacher (H1), we utilize test scores in reading literacy, mathematical literacy, and scientific literacy. Tests were constructed and evaluated by experts to ensure broad coverage of third and fourth grade curricula in the respective subjects (details: Bos *et al.*, 2003). In order to keep the test time to a reasonable length while at the same time ensuring a comprehensive coverage of the curricula, not all items were administered to all students. Instead, the test material was distributed among different booklet versions according to a multi-matrix design. Subsequently, within the framework of Item-Response-Theory, Maximum-Likelihood Estimates (MLE) were estimated which we employ as best estimates of the student performance (IPN, 2005). To test whether boys and girls are assessed differently (or behave differently) when being taught by a same-sex teacher (H2), we utilize teacher grades in German, Mathematics, and Science as dependent variables, controlling for the respective test scores. Grades run from 6 (insufficient/fail) to 1 (very good). Information about individual student grades was retrieved from the teacher questionnaires.

Independent Variables. The independent variables in our study are gender dummies for the teacher and the student. We are able to identify 2,434 male students (2,389 female students) with female German teachers and 553 male students (482 female students) with male German teachers. The share of female teachers in our sample ranges from 79% to 85%, which is why the number of teacher-student pairings varies somewhat across subjects (cf. Table 2). At the primary school level, the same teacher often teaches several subjects per class. Accordingly, in many cases one and the same teacher had to fill out several separate subject teacher questionnaires. In order not to be redundant, some teachers provided the standard socio-demographic information, including information on gender, in only one of the questionnaires. When socio-demographic information was missing, we imputed gender and other socio-demographic items when we could make sure we were dealing with the identical teacher.[10] In addition, we constructed a variable indicating whether a teacher was teaching all three subjects.

The remaining variables in this study represent controls for student, teacher, class, and school observables (cf. Table 1). To ensure that students' characteristics do not differ systematically between male and female teachers, we include at the student level parental socioeconomic status[11], students' migration background, number of children books at home, and students' age.[12] At the teacher/class level, we control for class size, share of male students in the class, and share of students with German language problems. For each teacher, we include age, number of years of teaching experience, a dummy indicating whether the teacher works part-time or full-time, and the amount of further training they engaged in while already working as a teacher.[13] External evidence (GFSO, 2009: own calculations) supports our data suggesting that female teachers are, on average, younger and less experienced and have a higher tendency to work part-time. Furthermore, female teachers tend to invest more in their further training (own auxiliary calculations based on IGLU-Data). We want to make sure that potential teacher-gender effects are not confounded by these factors. Next, to control for potential bias caused by the possibility that male and female teachers work at schools with different characteristics, we include school level controls. All school level information was provided through the school principal. We capture school resources by a dummy variable indicating whether or not a school library is available and by a factor score "shortage of teaching aids" based on a row of questions asking whether the lack of books, teaching materials, computers, audiovisual aids etc. constrains the school's capacity for providing adequate instruction. We also include a factor score "special needs" which indicates whether a school provides assistance for low-achieving students and students with language deficiencies, or whether it offers additional programs for high-achieving students. In addition, we include the percentage of students from economically disadvantaged backgrounds. The location of the school is controlled in terms of urbanity (urban, suburban, rural) and through dummies for East and West German states (*Länder*), between which the share of female teachers differs (GFSO, 2009).

We utilize multiple imputation to address missing data, which replaces missing values with several plausible values based on information observed in the sample (Allison, 2001; Rubin, 1987). Most missing values were found for parental socio-economic status, for which 25 percent of the sample had unknown values. Even in the face of a large amount of missing values, Rubin (1987: 114) states that only a few imputations are required to obtain estimates with a relatively high efficiency. We generate five datasets through imputation by chained equations (ICE), which has recently been implemented in Stata (Royston, 2004). All analyses are based on the imputed datasets, for which the results are combined to account for variation within and between imputed data sets.[14]

[Table 1 about here]

Results

We start out by documenting girls' and boys' performance scores and grades with same-sex and other-sex teachers across different subjects (Table 2). The share of female teachers differs across subjects (German: 83%, Mathematics: 78%, Science: 86%). Accordingly, the number of student-teacher-gender pairings differs across subjects as well.

[Table 2 about here]

The same-sex hypothesis states that students benefit from a same-sex teacher. Looking at girls, the raw data lends only weak support for this hypothesis. Girls with female teachers score between 8 and 16 points higher on performance tests than girls with male teachers, which equates roughly the winning margin of a ¼ school year (cf. Bos *et al.*, 2004: 56). However, the effects are not statistically significant. They also receive better (i.e. numerically lower) grades; however, the effect is only significant in Mathematics. Looking at boys, the hypothesis cannot prevail. Neither performance scores nor grades differ significantly between boys with male teachers and boys with female teachers. In addition, the (insignificant) direction of the teacher-gender effect is reversed, i.e. boys receive inferior grades from male teachers than from female teachers. One explanation could be, that male teachers grade rather strictly compared to their female teacher colleagues, independent of the student's gender. We tested this explicitly in a pooled dataset of girls and boys and found (insignificant) negative effects of male teachers across all subjects and outcomes (table not shown).[15] The direction of the effects suggests that there is no advantage of having a same-sex teacher for boys. If anything, male teachers have a consistent negative impact on both, boys and girls. However, such a negative male teacher effect (independent of student gender) is too weak to pass the test of significance.

In a next step, we run OLS-regressions of performance scores on teachers' gender separately for girls and boys (cf. specifications 1a and 1b, Table 3). We start with the empty model which naturally reproduces the descriptive figures from Table 2. The adjusted R-squared indicates that the teachers' gender explains virtually no variance in performance scores, neither for girls nor for boys. The same findings emerge when looking at grades (cf. specifications 5a and 5b, Table 4). Thus, across the three main subjects in primary school, and across objective performance measures (H1) as well as teachers' grades (H2), the 'same-sex teacher'-hypothesis has to be rejected.

However, these results could be quite misleading due to selection effects or because the hypotheses only hold in subgroups of the population. Selection effects could exist when students – or factors influencing students – differ systematically between male and female teachers, creating a spurious relationship between teacher gender and the outcome of interest. In the following section, we model a range of checks to verify whether our initial findings are robust. First, we try to tackle potential selection effects that usually pose a threat to the validity of empirical findings based on non-experimental data. Second, we try to estimate whether the initial absence of a teacher-gender effect is caused by a non-linear effect across different subgroups. For example, the gender of the teacher may have a significant effect only if students are exposed to the same teacher in all subjects.

Robustness Checks

Does selection mask a same-sex teacher effect?

First, suppose, for example, that male teachers are more likely to be assigned to boys with a propensity for lower achievement. In this case, the results in Table 2 could falsely suggest the absence of teacher-gender effects, because boys with a same-sex teacher would start out with a higher propensity for lower achievement, but would still benefit from their teachers. A comparison between test scores or grades of boys with same-sex teachers and boys with other-sex teachers would not reveal such teacher-gender effects.[16] Without longitudinal data, there is no direct way to address this classical concern of selection. However, we can control empirically whether male and female teachers are assigned to students with certain ascriptive traits that are usually confounded with high or low achievement, such as parents' socio-economic status, cultural capital in the home, migration background, or students' age.[17] If we hold such student traits constant and still find no evidence for a teacher-gender effect, we can most likely rule out the alternative explanation of biased student-teacher assignment.[18] We apply this strategy in specifications 2a (girls) and 2b (boys) for performance (Table 3) across several subjects. For the sake of clarity, we display only the main effect (that is the teacher gender effect).[19] As expected, the inclusion of student level controls increases the proportion of explained variance in the models drastically.[20] The teacher-gender effects, however, do not change by and large, and remain insignificant.

For our models with grades as dependent variables (testing H2), we first control for test scores (specifications 6a and 6b, Table 4) to see whether the only significant teacher gender effect in Mathematics for grade for girls (.13, cf. Table 2) can be explained by actual performance differences or by teacher's evaluation. As can be seen from the table, the teacher gender effect partially disappears, thus the slight difference in grading reflects the fact that girls with male teachers perform slightly less well than girls with female teachers, and are, net of performance differences, perceived more negatively. However, this effect is not significant. Next, we include student level controls (specifications 7a / 7b), just as in the performance model testing H1. There is very little change in coefficients across specifications; male teacher effects are small in magnitude, and none become significant.

Second, the bivariate findings in Table 2 could be biased by the presence of teacher traits (e.g. age) or classroom traits (e.g. class size) that are associated with a teacher's gender. We know from auxiliary analysis that female teachers are, on average, younger and therefore less experienced than their male counterparts. They also have a higher tendency to work part-time and are more likely to engage in further education while already filling a teaching position.[21] In specifications 3a/8a (girls) and 3b/8b (boys) we therefore control for observable teacher traits such as experience, amount of further education, employment status (part-time vs. full-time), and age. We also include a quadratic age term to capture non-linearity. In addition, we include the share of male students in class because we found that the share of male students in classes taught by a male teacher is slightly higher (53%) compared to classes taught by female teachers (48%). Again, results indicate that the teacher gender effect is not affected by these controls, neither in models where test scores are used as dependent variables (Table 3), nor in models where grades are used as dependent variables (Table 4). There is one exception: the effect of a male teacher for the reading performance of boys increases from -6.27 to -11.76 and becomes marginally significant. Note however, that the effect is negative and thus is inconsistent with a same-sex teacher benefit (i.e. boys should perform better with a male teacher). Given the weak level of significance one should be careful in interpreting this result. In any case there is no sign of a positive same-sex effect.

There might still be unobservable teacher traits (e.g. a teacher's socio-economic background) or class characteristics (e.g. classroom facilities) which confound our findings. Pooling the data on boys and girls, it is possible to control for all unobservable class and teacher traits by including class/teacher fixed effects (FE). Regarding the fixed effect regressions, we assume for the following structure:

$$y_a = x_a \beta^i_{g_a} + x_c \beta^c_{g_a} + g_a \gamma + g_t \delta_{g_a} + \varepsilon_c + \varepsilon_a \qquad (1)$$

y_a is the outcome of interest, x_a are covariates on the individual, and x_c on the class (including teacher) level. Furthermore g_a denotes student's gender, g_t teacher's gender (let $g_t = 1$ for male teachers) and ε_a is an idiosyncratic (unobserved) error term that is assumed to be uncorrelated with all covariates. ε_c denotes (unobserved) class fixed effects - potentially correlated with the observed covariates. We identify our parameters by within class differences between students a and b:

$$\Delta y_{a,b} = y_a - y_b = \begin{pmatrix} x_i \\ x_c \\ g_i \\ g_t \end{pmatrix}' \begin{pmatrix} \beta^i_{g_a} \\ \beta^c_{g_a} \\ \gamma \\ \delta_{g_a} \end{pmatrix} + \varepsilon_c + \varepsilon_a - \begin{pmatrix} x_b \\ x_c \\ g_b \\ g_t \end{pmatrix}' \begin{pmatrix} \beta^i_{g_b} \\ \beta^c_{g_b} \\ \gamma \\ \delta_{g_b} \end{pmatrix} - \varepsilon_c - \varepsilon_b$$

$$= \begin{pmatrix} x_c \\ g_t \end{pmatrix}' \begin{pmatrix} \beta^c_{g_a} - \beta^c_{g_b} \\ \delta_{g_a} - \delta_{g_b} \end{pmatrix} + (g_a - g_b)\gamma + x_a \beta^i_{g_a} - x_b \beta^i_{g_b} + (\varepsilon_a - \varepsilon_b) \qquad (2)$$

From this last equation, we see that the class fixed effects cancel out. Additionally, coefficients of variables not varying within class are no longer identified. As a result, the effect of teacher gender itself cannot be identified any more, as it is part of the class FE. However, an interaction effect, i.e. the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect), is still identified under weaker assumptions than before.[22] Now, if there actually were a positive same-sex effect, the difference between a male teacher effect on boys and girls should be large and positive.

FE-Models for all subjects are presented in the last specification of Tables 3 and 4. There are 308 unique teachers/classes per subject associated with 5,858 first-differenced observations. Results support what we observed so far. All effects have the same sign and magnitude as indicated by the differences between specifications (3b)-(3a) and (8b)-(8a), as summarized in Table II (Appendix). Interactions from the FE regressions are less than one standard deviation from what is indicated by standard OLS. Moreover, they are far from being significant except for Science where they pass the test at the 10% level.[23]

[Tables 3 and 4 about here]

Finally, a bias could occur if school characteristics such as school resources or the urbanity of a school location is associated with a teacher's gender. For example, it is conceivable that male teachers are more likely to find employment at less favourable schools. If, in addition, boys happen to be especially prone to suffer from unfavourable school conditions, the male teacher effect for boys would be biased downwards and a same-sex teacher effect could be masked. It needs to be emphasized that we do not expect such biases because, as mentioned above, the application procedure of teachers to schools used to be very indirect. Auxiliary analysis confirms the expectation that the likelihood of finding a male teacher assigned to a school with certain observable school characteristics (such as different kinds of resources, school location, and student composition) does not differ systematically by observable school characteristics (cf. Table III in Appendix). However, the likelihood of finding a male teacher does differ across the German states. This is because the share of female teachers is higher in former East Germany (GFSO, 2009), which is a relict of the higher female labour market participation in the German Democratic Republic. The inclusion of a dummy to account for East/West differences, however, does not change our findings.

Do subgroup differences mask a same-sex teacher effect?

So far, the only finding which is striking is the extent to which the match between teacher gender and student gender is unimportant. However, before negating a phenomenon many scholars and politicians believe in, we run an additional set of tests before conceding the case for a no same-sex teacher effect. An effect of the teachers' gender may become relevant, if the 'treatment' (that is the exposure to the same teacher) is stronger. In Germany, each primary school class usually has one teacher, who teaches most major subjects in this class and who stays with the same class for two or more years. The idea is to have a stable psychological parent for the students, who knows the students very well and is thus able to respond more closely to their personal difficulties. The downside is that potential difficulties, such as negative teacher-gender effects, are more prevalent and pertain to all subjects. We ran the full set of regressions on the subgroup of students who were taught by the same teacher in all subjects under investigation, i.e. German, Mathematics, and Science (49 percent of the sample). Table 5 displays our findings. Again, we must conclude that a teacher-gender effect is not detectable. There is only one marginally significant effect of having a male teacher for boys in mathematics when looking at grades. The direction of the effect, however, speaks against the same-sex teacher hypothesis.

Next, we ran the same set of regressions on students who were most likely taught by the same teacher for four years (i.e. we excluded students from schools for which the principle stated that students are usually taught by the same teacher for only 3 years or less). For this subgroup (56 percent of the sample), a negative effect of having a male teacher becomes evident when looking at German test scores (cf. Table 5). But this effect is around -21 points for both boys and girls, which speaks once more against a same-sex teacher advantage (with our coding one would expect a positive effect for boys).

In accordance with theoretical considerations in the same-sex literature, we hypothesised in H2 that a mismatch between the habitus of boys and the habitus of female teachers might be responsible for the academic underachievement of boys. It might be argued that this mismatch should be stronger for boys with a particularly pronounced male habitus. Since boys from migrant and low-SES backgrounds are commonly assumed to have a particularly pronounced male habitus (Connell, 1987; Connell, 1995; King, 2005), these subgroups should be especially harmed by female teachers. We tested this hypothesis by running regressions on the subsample of students from low-SES backgrounds, i.e. the 25th quantile of the socio-economic status distribution as well as on students with a migration background (cf. Table 5). Once again, results suggest a firm decline of our hypothesis. Across all outcomes (i.e. performance scores and grades in German, Mathematics, and Science), male teacher effects are insignificant for boys and for girls in both subgroups.

[Table 5 about here]

Summary and Conclusion

In past decades, school systems in nearly all countries of the world have seen a rather unanticipated gender-specific change in educational achievement, with females gaining advantages over males. At the same time, the share of female teachers has grown in almost all countries of the western world. In this context, we examine the empirical evidence regarding one question increasingly addressed in educational research and politics alike: Can the growing “feminization” of the teacher profession explain the emerging disadvantage for boys in educational attainment? The claim that a lack of male teachers is responsible for the ‘boy crisis’ in education is highly popular internationally, both in some parts of academia and among the broader public. This is astonishing, given that there is hardly any solid empirical evidence to support the same-sex hypothesis.

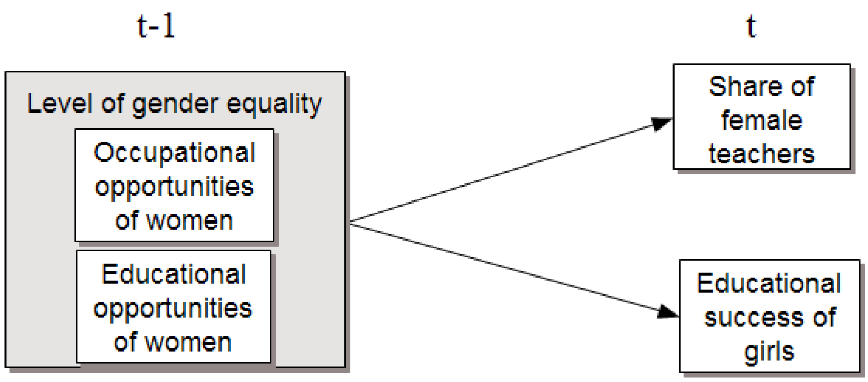
After evaluating the evidence on this research question, we isolate a teacher’s gender effect by exploiting two features of the primary German educational system: a long ‘exposure time’ of students to the same teacher (at least two years) and a quasi-random assignment of students to teachers. We use large-scale data from IGLU-E/PIRLS and focus on different scholastic outcomes (‘gender-blind’ test scores and more subjective teacher’s grades) across different subjects (German, Mathematics, and Science). We find that boys do not benefit from male teachers and girls do not - at least not significantly - benefit from female teachers, neither with regard to their academic performance as measured by test scores, nor with regard to their grades. This non-existence of a teacher-gender effect is evident across the three major subjects in German primary school, namely German (a typical ‘female’ domain), Math (a typical ‘male’ domain), and Science (which has to be considered rather ‘gender neutral’ in the German primary school curriculum (cf. footnote 6)). In summary, we can say that the “feminization” of (primary) schooling has neither led to a weaker development of the academic skills of boys, nor has it improved the academic skills of girls. Grading has also been found to be independent of teacher gender. Children from migrant and low-SES families do not benefit academically from being taught by a same-sex teacher; neither do children who are taught by the same teacher in all subjects. In other words, the call for more male teachers to address boys' educational and psycho-social needs in school is not based on sound empirical evidence, at least not with regard to primary schools. Furthermore, our results suggest that the unqualified call for more male teachers may have unintended consequences. Both boys and girls who were taught by a male German teacher for 4 years (56 percent of the sample) had significantly less well developed reading skills than students who were taught by a female German teacher for four years. More research is needed on why this finding might be and in how far it can be generalized. At this point, we can only speculate why male German teachers who have taught a class for 4 years should have a negative effect on all students. It is conceivable that female teachers may have higher reading skills themselves, which they are better able to convey to their students. To our knowledge, studies on the gender-specific reading skills of German teachers do not exist, however. Nevertheless, calling for more male teachers involves the danger of unintentionally harming the reading skills of both boys and girls by having them taught by male teachers.

What is most striking about our findings is the extent to which the match between teacher gender and student gender is irrelevant.[24] In this context, it is important to carefully consider the degree to which our findings may be generalized. First, there might be teacher gender effects at a later student age, i.e. during adolescence. This may help to explain why Dee (2007) found a correlation between the academic achievement of US secondary school students and same-sex teachers. Second, it is possible that boys who grow up without a male role model at home do benefit from having male teachers. This hypothesis could not be tested with the data available.

One additional aspect of this paper seems to require further explanation, however. Why is it that the correlation between the share of female teachers and gender-specific educational achievement, as clearly shown to exist at the country level in Figure 1, cannot be supported with empirical evidence at the individual level? We would like to argue that this puzzle results from a classical ecological fallacy. To find out about this, we need to ask how cross-country differences regarding the percentage of female teachers came about in the first place, and why the total share of female teachers has risen over past decades. Two aspects of the teaching profession become evident when comparing it across countries: Firstly, completing higher education is a prerequisite for gaining teacher licensure in most countries. Secondly, teaching is a female sex-segregated profession in literally all advanced societies, while it is male dominated in developing countries (most visibly in African countries, cf. figure 1). In the latter countries, women do not obtain higher education and are consequently not able to work as teachers – thus the share of female teachers is low. In addition, recent research has shown, that once the level of female tertiary enrolment begins to rise in a given country, having a higher percentage of women in female sex-segregated academic professions is an almost automatic consequence, as long as female preferences for traditionally female occupations do not change (Charles and Bradley, 2009). The argument is that in advanced industrial societies, it is more 'culturally legitimate' to pursue self-realization. As such, men and women are free to express their 'gendered selves', when choosing a field of study and an occupation. As a consequence, the gender-segregation across academic fields (such as education) and subsequently across occupations (such as teaching) intensifies. This means that the current share of female teachers at school may be deduced from women's educational opportunities in past decades and their opportunities of participating in the labour market today - two indicators of general gender equality (figure 2).

But why are girls and women today academically more successful in countries, in which the gender equality has been higher in the past? First of all, it is conceivable that socio-cultural and institutional factors that have helped women in past decades to become successful in the academic labour market are beneficial for girls and women today. For example, a declining discrimination in the labour market (possibly because 'female' qualities are appreciated more in societies with a rising service sector and a decreasing industrial sector) may give rise to girls' and women's aspirations. As girls and women think about their expectations to succeed in education, the prevailing labour market opportunities for women in their country influences whether they expect to complete an academic degree and obtain an academic job position (cf. McDaniel, 2010). This argument is closely related to the expectation, that in countries where women have experienced relatively higher gender equality in the past, girls have more role models who demonstrate: females can be successful in the labour market if they are successful in the education system (Mickelson, 1989). This may very likely encourage females to appreciate educational outcomes (Buchmann and DiPrete, 2006) and utilize their academic potential. Evidence of the fact that female employment levels and general gender equity have an impact on gender-specific educational achievement could be provided by several authors (Baker and Jones, 1993; Else-Quest *et al.*, 2010; Guiso *et al.*, 2008; McDaniel, 2010; Riegle-Crumb, 2005). The so called “gender-stratification hypothesis” posits, that gender differences in some educational outputs, like mathematic competencies, are closely related to cultural variations in opportunity structures for girls and women.

Figure 2: Relationship of gender equality with the share of female teachers and the educational success of girls



Source: own illustration

In sum, not the gender of the teacher is relevant for the increasing educational success of girls. Rather, equal gender opportunities in a given country encourage girls to realize their academic potential, while at the same time making it more likely for women to become teachers.

References

ABC-News (2004). Scholarships to target declining male teacher numbers. <www.abc.net.au/news/stories/2004/05/03/1100090.htm> [accessed 8 July 2010].

Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.

Ammermüller, A. and Dolton, P. (2006). Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA. *ZEW Discussion Paper* 06-600.

Bacher, J., Beham, M. and Lachmayr, N. (2008). *Geschlechterunterschiede in der Bildungswahl*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baker, D. P. and Jones, D. P. (1993). Creating gender equality: cross-national gender stratification and mathematical performance. *Sociology of Education* **66**, 91-103.

BBC-News (2000). Male teachers for 'role models'. <http://news.bbc.co.uk/2/hi/uk_news/ education/893313.stm> [accessed 8 July 2010].

Blossfeld, H. P., Bos, W., Hannover, B., Lenzen, D., Müller-Böling, D., Prenzel, M., Wößmann, L. (2009). Geschlechtsdifferenzen im Bildungssystem. Jahresgutachten 2009 Aktionsrat Bildung.

Blount, J. (2000). Spinsters, bachelors and other gender transgressors in school employment, 1850-1990. *Review of Educational Research* **70**, 83-101

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. and Walther, G. (2004). *IGLU: Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. and Walther, G. (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.

Brandes, H. (2002). *Der männliche Habitus*, Vol. 2. Opladen: Leske+Budrich.

Buchmann, C. and DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review* **71**, 515-541.

Budde, J. (2008). Bildungs(miss)erfolge von Jungen und Berufswahlverhalten bei Jungen / männlichen Jugendlichen, Vol. 23. Bildungsforschung Edition: BMBF, Referat Chancengerechtigkeit in Bildung und Forschung.

Carter, R. S. (1952). How invalid are marks assigned by teachers. *Journal of Educational Psychology* **43**, 218-228.

Charles, M. and Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology* **114**, 924-976.

Cole, N. S. (1997). *The ETS Gender Study: How Females and Males Perform in Educational Settings.* Princeton, NJ: Educational Testing Service.

Connell, R. W. (1995). *Masculinities*. Berkeley: University of California Press.

Connell, R. W. (1987). *Gender and power. Society, the Person, and Sexual Politics.* Stanford: Stanford University Press.

Dammasch, F. (2007). *Jungen in der Krise: Das schwache Geschlecht? Psychoanalytische Überlegungen*. Frankfurt am Main: Brandes and Apsel.

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

Diefenbach, H. and Klein, M. (2002). Bringing boys back in. *Zeitschrift für Pädagogik* **48**, 938-958.

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

Eagly, A. H. and Chrvala, C. (1986). Sex differences in conformity: status and gender role interpretations. *Psychology of Woman Quarterly* **10**, 203-220.

Ehrenberger, R. G., Goldhaber, D. D. and Brewer, D. J. (1995). Do teachers' race, gender and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review* **48**, 547-561.

Else-Quest, N. M., Hyde, J. S. and Linn, M. C. (2010). Cross-national patterns of gender differences in Mathematics: a meta-analysis. *Psychological Bulletin* **136**, 103-27.

Eurostat (2009). *Eurostat Database*. European Commission. <http://epp.eurostat.ec.europa.eu/portal/page/portal/education/data/database> [accessed 8 July 2010].

Ganzeboom, H. B. G., De Graaf, P. M., Treiman, D. J. and De Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research* **21**, 1-56.

German PISA Consortium (2007). *PISA. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster/ New York/ München/ Berlin: Waxmann.

GFSO (2009). Fachserie 11.1 - Allgemeinbildende Schulen Schuljahr 2008/2009, *German Federal Statistic Office*.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology* **60**, 549-576.

Guiso, L., Monte, F., Sapienza, P. and Zingales, L. (2008). Culture, gender, and math. *Science* **320**, 1164-1165.

Hannover, B. (2008). Vom biologischen zum psychologischen Geschlecht: Die Entwicklung von Geschlechtsunterschieden. In Renkl, A. (Ed.), *Lehrbuch der pädagogischen Psychologie*. Bern: Huber, 339-388.

Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of Economics of Education* **2**, 1053-1078.

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

Helbig, M. (2010). Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **62**, 93-111.

Holmund, H. and Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics* **15***,* 37-53.

Hopf, D. and Hatzichristou, C. (1999). Teacher gender-related influences in greek schools. *British Journal of Education Psychology* **69**, 1-18.

IES (2009). *NAEP Data Explorer: National Center for Education Statistic.* <http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx> [accessed 8 July 2010].

IPN (2005). *Research Area 5 - Educational Monitoring and Research Methodology.* Research Report 2003/2004.

King, V. (2005). Bildungskarrieren und Männlichkeitsentwürfe bei Adoleszenten aus Migrantenfamilien. In: King, V. and Flaake, K. (Eds.), *Männliche Adoleszenz. Sozialisation und Bildungsprozesse zwischen Kindheit und Erwachsensein.* Frankfurt/New York: Campus, 57-76.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex role concepts and attitudes. In MacCoby, E. E. (Ed.), *The Development of Sex Differences.* Stanford: Stanford Univ. Press, 82-172.

Marsh, H. W., Martin, A. J. and Cheng, J. H. S. (2008). A multilevel perspective on gender in classroom motivation and climate: potential benefits of male teachers for boys? *Journal of Educational Psychology* **100**, 78-95.

Martino, W. J. (2008). Male teachers as role models: addressing issues of masculinity, pedagogy and the re-masculinisation of schooling. *Curriculum Inquiry* **38**, 189-223.

McDaniel, A. (2010). Cross-national gender gaps in educational expectations: the influence of national-level gender ideology and educational systems. *Comparative Education Review* **54**, 27-50.

Mickelson, R. A. (1989). Why does Jane read and write so well? The anomaly of women's achievement. *Sociology of Education* **62**, 47-63.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. and Kennedy, A. M. (2003). *PIRLS 2001 International report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M. and Flaherty, C. L. (Eds.) (2002). *PIRLS 2001 Encyclopedia*. Boston: PIRLS International Study Center.

OECD (2009). OECD.Stat Extracts: Organisation for Economic Co-Operation and Development.

Pollack, W. S. (2006). The "war" for boys: hearing "real boys" voices, healing their pain. *Profesional Psychology - Research and practice* **37**, 190-195.

Powell, B. and Downey, D. B. (1997). Living in single-parent households: an investigation of the same-sex hypothesis. *American Sociological Review* **62**, 521-539.

Quenzel, G. and Hurrelmann, K. (2010). Geschlecht und Schulerfolg: Ein soziales Stratifikationsmuster kehrt sich um. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **62**, 61-91.

Riegle-Crumb, C. (2005). The cross-national context of the gender gap in math and science. In Hedges, L. V. and Schneider, B. (Eds.), *The Social Organization of Schooling*. New York: Russell Sage Foundation, 227-243.

Royston, P. (2004). Multiple imputation of missing values. Stata Journal **4**, 227-241.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rury, J. L. (1989). Who became teachers? The social characteristics of teachers in american history. In Warren, D. (Ed.), *American Teachers: Histories of a Profession at Work.* New York: Macmillan, 9-48.

Schwippert, K. (2007). *The impact of PIRLS in Germany*. In Schwippert, K. (Ed.), *Progress in Reading Literacy: The Impact of PIRLS 2001 in 13 Countries.* Münster: Waxmann, 93-108.

Sexton, P. (1969). *The Feminized Male: Classroom White collars, and the Decline of Manliness*. New York: Random House.

Sokal, L., Katz, H., Chaszewski, L. and Wojcik, C. (2007). Good-bye, Mr. Chips: male teacher shortages and boys' reading achievement. *Sex Roles* **56**, 651-659.

Sommers, C. H. (2000). *The War against Boys: How Misguided Feminism is Harming our Young Men*. New York: Touchstone.

Stamm, M. (2008). Underachievement von Jungen: Perspektiven eines internationalen Diskurses. *Zeitschrift für Erziehungswissenschaft* **11**, 106-124.

Steele, C. M. (1997). A threat in the air: How stereotypes shape the intellectual identities and performance of women and African-Americans. *American Psychologist* **52**, 613-629.

TDA (2009) Wanted: male primary school teachers <www.tda.gov.uk/about/mediarelations/2009/200909.aspx> [accessed 8 July 2010].

Treptow, E. and Rothland, M. (2005). „Jeder Schule ihre Lehrer“? - Empirische Befunde zum Auswahlverfahren bei der Besetzung von Lehrerstellen in Nordrhein-Westfalen. *Zeitschrift für Erziehungswissenschaft* **8**, 305-320.

UNESCO (2009). *Global Education Digest 2009: Comparing Education Statistics Across the World*. Montreal: UNESCO Institute for Statistic.

UNESCO (2010). Unesco Institute for Statistic - Key statistical tables for education. <http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx?IF_ActivePath=P,50&IF_Language=eng> [accessed 08 July 2008].

Wagemaker, H., Taube, K., Munck, I., Kontogiannopoulou-Polydorides, G. and Martin, M. (1996). *Are Girls Better Readers?: Gender Differences in Reading Literacy in 32 Countries*. Amsterdam: IEA.

Walther, G., Schwippert, K., Lankes, E.-M. and Stubbe, T. C. (2008). Können Mädchen doch rechnen? Vertiefende Analysen zu Geschlechtsdifferenzen im Bereich Mathematik auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU. *Zeitschrift für Erziehungswissenschaft* **11**, 30-46.

Willingham, W. W. and Cole, N. S. (1997). *Gender and Fair Assessment*. Mahwah: Lawrence Erlbaum.

Notes

1 Of course, this correlation can be spurious. Cf. ‘Summary and Conclusion’ for a discussion on this.

2 The PIRLS study was conducted in 2001. 35 countries participated in the study, which assessed the reading literacy of fourth-graders. The 2001 German supplemental sample collected additional data on students’ math and science skills as well as on their grades in German, Mathematics, and Science (Sachkunde).

3 In addition, grades are a useful component because of their relevance for students’ transition into one of Germany’s distinct secondary school types, which in large parts determines a student’s future educational and occupational career.

4 Science (Sachunterricht) is a subject concerned with social, historical, scientific and geographical topics. The Sachunterricht curriculum suggests that this is neither a clear “boys’ domain” nor a clear “girls’ domain” but rather “gender neutral”.

5 This so-called *Listeneinstellungsverfahren* was changed in recent years in favor of a procedure that grants more agency and autonomy to schools (*schulscharfes Verfahren*). These reforms do not pertain to our 2001 data set, however, since the teachers in question were appointed between the 1960s and the 1990s.

6 Helbig (2010) analysed the effect of the ratio of male teachers at the school level on boys' and girls' grades. He found that an increase of the male teacher ratio in a school by 10 percent would raise the grades of boys in mathematics by less than 0.1 grade points. There were no effects of the ratio of male teachers on mathematics grades for girls or on German grades for both, boys and girls.

7 Access to the IGLU-E datasets has been kindly provided by the Research Data Centre at the Institute for Educational Progress (Institut zur Qualitaetsentwicklung im Bildungswesen, IQB).

8 Note that the German sample for the international comparison consists of 8,997 students in 211 schools from all 16 states. However, the states of Brandenburg, Mecklenburg-West Pomerania, Lower Saxony, and Saxony-Anhalt did not participate in the extension and hence do not provide information on Mathematics and Science, amongst other background information. We see no reason, why results might be biased by the exclusion of these 4 states, as they are very similar to the included 12 states in terms of school system characteristics, student performance and demographic composition including the feminisation of the teacher workforce.

9 The fact that the participation rate is somewhat lower than the international average is probably caused by the necessity of obtaining written parental consent.

10 We used a range of socio-demographic, teaching, and school related variables to identify identical teachers across the different teacher questionnaires.

11 Socio-economic status is measured by the Highest International Socio-economic Index of Occupational Status (HISEI), which corresponds to the highest occupational index score of the student's father or mother (Ganzeboom *et al.*, 1992).

12 We also included a quadratic age term to capture non-linearities. The same procedure has been applied to teachers' age.

13 German teachers were asked how often they read material about instruction or didactics. Mathematics and science teachers were asked how many times within the last two years they participated in further education seminars in their subjects.

14 It has been shown that this procedure has advantages over listwise deletion in terms of obtaining unbiased estimators. For a recent overview see Graham (2009).

15 Analysis can be obtained from the authors upon request. Another explanation could be that students do not learn as much from male teachers. We test this in the remainder of the study.

16 The same argument may be applied to girls and female teachers, of course.

17 In specifications not reported here, we also included intelligence test scores, which are supposed to be time-invariant. However, the time-invariance of intelligence assessment is controversial in psychological research and could be influenced by the teacher's gender as well. We therefore do not include intelligence test scores in the models shown. The inclusion of intelligence test scores does not change results regarding the teacher-gender effect.

18 Selection on unobservable characteristics would in principle still be possible, but given the wide range of covariates available to us, we consider this rather unlikely.

19 The full set of covariates can be seen in Table I (appendix), exemplarily for reading test score as an outcome.

20 Interestingly, the adjusted R-squared increases most in models with reading literacy. This is because our control variables ‚books at home' and migration background (which is highly correlated with language proficiency) are better predictors of reading literacy than of the other performance domains.

21 They do not differ according to their formal training, as all teachers have to obtain a teaching degree in order to be eligible for teaching.

22 The sum of the same-sex teacher effects for boys and girls is equivalent to the difference $\delta_{g_a} - \delta_{g_b}$ from equation (2). This result can be obtained by just defining g_t relative to g_a instead of as always one for male teachers. Also note that in our specification, for each subject, we have one teacher per class. In this respect, teacher and class fixed effects are identical.

23 In the case of test scores, the loss of degrees of freedom is not sufficiently compensated by higher explanatory power of the model, thus leading to a lower adjusted R-squared.

24 This is not to say that teacher effects do not exist. In fact, many studies show that a teacher explains large shares of the variance in student outcomes (Hanushek and Rivkin, 2006). However, the important factor is probably the quality of teaching rather than the gender of the person delivering it (Marsh *et al.*, 2008).

Tables

Table 1: *Descriptive statistics (averaged over multiply imputed data sets, weights apply)*

		Girls (n=2,871)			Boys (n=2,987)		
		Mean	SD	Min/Max	Mean	SD	Min/Max
Performance [i]							
German		511.9	97.9	-83.4/ 896.8	498.4	95.9	-78.0/ 904.8
Mathematics		494.6	98.3	48.1/ 916.6	511.5	98.5	42.1/ 916.6
Science		492.9	99.7	-26.0/ 844.5	510.6	98.9	124.0/ 844.6
Grades							
German		2.59	.93	1/6	2.94	.89	1/6
Mathematics		2.80	.99	1/6	2.69	.96	1/6
Science		2.48	.90	1/6	2.53	.87	1/6
% male teachers							
German		.16	.37	0/1	.17	.38	0/1
Mathematics		.22	.41	0/1	.22	.41	0/1
Science		.15	.36	0/1	.15	.36	0/1
Student level controls							
Migration background		.25		0/1	.25		0/1
parental hisei		49.59	14.99	16/90	49.78	15.19	16/90
age [ii]		10.48	.47	8.92/13.08	10.58	.52	8.5/13.25
books [iii]	(0-10)	5.77			8.29		
	(11-25)	17.43			20.09		
	(26-50)	33.00			32.31		
	(51-100)	27.21			23.80		
	(more than 100)	16.59			15.51		
Class level controls							
Size		22.86	3.92	9/32	22.58	4.02	9/32
% male students		.48	.11	.15/.88	.53	.11	.15/1
% language problems		.06	.12	0/1	.06	.13	0/1
German							
teacher age [ii]		48.25	9.71	26/63	48.62	9.66	26/63
teacher experience		23.26	11.59	1/42	23.80	11.59	1/42
teacher further training [iiii]		2.45	.60	1/4	2.47	.60	1/4
% part-time		.38	.49	0/1	.38	.49	0/1
Mathematics							
teacher age		47.79	9.66	26/64	47.92	9.60	26/64
teacher experience		22.74	11.31	0/42	22.91	11.35	0/42
teacher further training	(>5)	6.14			4.74		
	(3-4)	14.71			15.12		
	(1-2)	39.48			39.04		
	(0)	39.67			41.10		
% part-time		.30	.46	0/1	.28	.45	0/1
Science							
teacher age		47.42	10.23	25/63	47.67	10.14	25/63
teacher experience		22.40	12.04	1/42	22.82	11.99	1/42
teacher further training	(>5)	3.73			3.22		
	(3-4)	8.89			10.96		
	(1-2)	31.44			28.89		
	(0)	55.94			56.93		
% part-time		.34	.47	0/1	.33	.47	0/1
School level controls							
school library available		.51	.50	0/1	.50	.50	0/1
special needs		.46	.29	0/1	.46	.30	0/1
shortage of teaching aids		2.02	.63	.29/4	2.03	.66	.29/4
East Germany		.09	.28	0/1	.09	.29	0/1
% students econ. disadvantaged		49.07			49.80		
backgrounds	(0-10)						
	(11-25)	31.36			29.84		
	(26-50)	10.79			12.61		
	(>50)	8.78			7.75		
school location	(urban)	32.27			30.46		
	(suburban)	23.75			23.05		
	(rural)	43.98			46.49		

[i] Performance scores vary up to very high or sometimes negative values. This is due to the maximum-likelihood-estimates, which are estimated based on a rotated test design within the framework of item-response-theory. This should not affect regression outcomes compared to a score normalized to some tighter bounds.
[ii] we also include a quadratic age term in the models to capture non-linearity.
[iii] we display percentages for all categorical variables.
[iiii] cf. footnote 12.

Table 2: *Students' performance and grades by student and teacher gender (averaged over multiply imputed data sets, weights apply)*

	Girls			Boys		
	Same-sex teacher	Other-sex teacher	Diff.	Same-sex teacher	Other-sex teacher	Diff.
Performance Score						
German	513	505	8	498	498	0
Mathematics	496	488	8	515	510	5
Science	495	479	16	510	510	0
Grade						
German	2.58	2.69	-.11	2.99	2.92	.07
Mathematics	2.77	2.90	-.13 **	2.72	2.68	.04
Science	2.48	2.53	-.05	2.62	2.52	.10
Sample size [i]						
German	2389	482		553	2434	
Mathematics	2252	619		650	2337	
Science	2486	385		438	2549	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$ (two-tailed T-Test, standard errors adjusted for class-level clustering).

[i] The sample sizes differ slightly across imputations because some of the teacher's gender has been imputed. We display the case numbers of the first imputation in the tables. All analyses, however, are based on all 5 imputations according to Rubin's (1987) rule.

Table 3: Estimated effect of a male teacher on test scores for different subjects by student gender (OLS & FE Regressions)

	Girls OLS			*Boys OLS*			*Pooled Boys&Girls FE*[i]
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)	(4)
	Reading			Reading			Reading
Male teacher	-7.65	-9.42	-5.69	-0.36	-6.27	-11.76*	-2.87
	(8.03)	(6.30)	(7.31)	(7.65)	(5.73)	(6.84)	(6.55)
Student level controls	-	x	x	-	x	x	x
Class/teacher level controls	-	-	x	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	x
Adjusted R-squared	0.001	0.219	0.228	0.000	0.198	0.204	0.167
	Math			Math			Math
Male teacher	-8.38	-7.01	-4.27	5.49	3.73	2.46	8.69
	(6.70)	(4.85)	(5.48)	(6.44)	(5.25)	(6.01)	(6.64)
Student level controls	-	x	x	-	x	x	x
Class/teacher level controls	-	-	x	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	x
Adjusted R-squared	0.001	0.131	0.140	0.000	0.124	0.127	0.099
	Science			Science			Science
Male teacher	-16.28	-12.03	-8.11	-0.49	1.38	-0.66	15.52*
	(10.64)	(7.72)	(8.81)	(7.98)	(6.05)	(6.95)	(9.33)
Student level controls	-	x	x	-	x	x	x
Class/teacher level controls	-	-	x	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	x
Adjusted R-squared	0.004	0.164	0.170	0.000	0.148	0.149	0.124
Observations	2871	2871	2871	2987	2987	2987	5858
Number of classes	308	308	308	308	308	308	308

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply. FE Regressions allow for differing coefficients of the control variables by student's gender, to reach the same flexibility as the separate gender models.
[i] The interaction effects denote the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect).

Table 4: Estimated effect of a male teacher on teachers' grades for different subjects by student gender (OLS & FE Regressions)

	Girls				***Boys***				***Pooled Boys&Girls FE*** [i]
	(5a)	(6a)	(7a)	(8a)	(5b)	(6b)	(7b)	(8b)	(9)
	German				German				German
Male teacher	0.11	0.06	0.08	0.02	0.07	0.07	0.11	0.03	0.01
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)
Reading literacy test score	-	x	x	x	-	x	x	x	x
Student level controls	-	-	x	x	-	-	x	x	x
Class/teacher level controls	-	-	-	x	-	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	-	-	x
Adjusted R-squared	0.002	0.340	0.398	0.406	0.001	0.295	0.358	0.372	0.453
	Math				Math				Math
Male teacher	0.13**	0.08	0.08	0.06	0.04	0.07	0.09	0.04	-0.03
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Mathematical literacy test score	-	x	x	x	-	x	x	x	x
Student level controls	-	-	x	x	-	-	x	x	x
Class/teacher level controls	-	-	-	x	-	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	-	-	x
Adjusted R-squared	0.003	0.307	0.365	0.374	0.000	0.282	0.317	0.325	0.374
	Science				Science				Science
Male teacher	0.05	-0.02	-0.02	-0.03	0.10	0.09	0.10	0.10	0.15*
	(0.14)	(0.12)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Scientific literacy test score	-	x	x	x	-	x	x	x	x
Student level controls	-	-	x	x	-	-	x	x	x
Class/teacher level controls	-	-	-	x	-	-	-	x	x
Class/teacher fixed effects	-	-	-	-	-	-	-	-	x
Adjusted R-squared	0.002	0.204	0.302	0.322	0.004	0.181	0.265	0.285	0.320
Observations	2871	2871	2871	2871	2987	2987	2987	2987	5858
Number of classes	308	308	308	308	308	308	308	308	308

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply.
FE Regressions allow for differing coefficients of the control variables by student's gender, to reach the same flexibility as the separate gender models.
[i] The interaction effects denote the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect).

Table 5: Estimated effect of a male teacher on students' test scores and grades for different subgroups (OLS Regressions)

	Girls				***Boys***			
	same teacher all subjects	same teacher 4 years	migrants	.25 low SES quantile	same teacher all subjects	same teacher 4 years	migrants	.25 low SES quantile
	Test scores							
	German				German			
Male teacher	-7.80	-20.24*	-11.94	1.51	-8.36	-21.44**	-5.54	-8.12
	(7.54)	(10.98)	(12.86)	(17.05)	(8.14)	(9.66)	(16.07)	(13.69)
Adj./Pseudo R^2	0.244	0.207	0.179	0.174	0.216	0.201	0.190	0.116
	Math				Math			
Male teacher	-10.72	10.06	12.19	14.24	1.88	-0.74	-3.67	-2.22
	(7.28)	(7.91)	(12.56)	(9.49)	(7.69)	(8.41)	(9.97)	(11.25)
Adj./Pseudo R^2	0.154	0.134	0.143	0.129	0.121	0.125	0.108	0.083
	Science				Science			
Male teacher	-11.65	-9.72	-12.20	-7.85	1.40	-3.88	-4.91	-4.70
	(10.13)	(13.54)	(16.27)	(16.99)	(7.52)	(10.62)	(14.09)	(12.61)
Adj./Pseudo R^2	0.171	0.174	0.164	0.163	0.165	0.144	0.138	0.096
	Grades							
	German				German			
Male teacher	0.08	-0.16	0.17	0.12	0.11	-0.07	0.10	0.06
	(0.09)	(0.12)	(0.14)	(0.13)	(0.10)	(0.13)	(0.14)	(0.12)
Adj./Pseudo R^2	0.424	0.411	0.356	0.365	0.359	0.386	0.398	0.352
	Math				Math			
Male teacher	0.06	0.04	0.15	0.15	0.14*	-0.09	0.02	0.05
	(0.08)	(0.07)	(0.12)	(0.13)	(0.08)	(0.09)	(0.10)	(0.11)
Adj./Pseudo R^2	0.373	0.366	0.315	0.306	0.335	0.340	0.306	0.334
	Science				Science			
Male teacher	0.09	-0.25**	0.09	-0.03	0.16	-0.04	0.09	-0.02
	(0.09)	(0.12)	(0.16)	(0.15)	(0.10)	(0.13)	(0.19)	(0.14)
Adj./Pseudo R^2	0.279	0.357	0.267	0.283	0.272	0.300	0.293	0.298
Observations	1414	1611	754	691	1464	1666	805	748

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply. All models include student level and class/teacher level controls.

Appendix

Table I: Effect of a male teacher on girls and boys reading test scores (OLS Regressions), *all control variables shown*

	Girls – reading test score			Boys – reading test score		
	(1a)	(2a)	(3a)	(1b)	(2b)	(3b)
Male teacher	-7.65	-9.42	-5.69	-0.36	-6.27	-11.76*
	(8.03)	(6.30)	(7.31)	(7.65)	(5.73)	(6.84)
Student level controls						
age		426.50***	427.80***		347.00***	321.50***
		(101.87)	(105.10)		(102.77)	(97.45)
stud_age2		-20.67***	-20.71***		-17.00***	-15.75***
		(4.78)	(4.94)		(4.82)	(4.54)
books (0-10)		-37.30***	-35.59***		-38.78***	-37.28***
		(11.40)	(11.16)		(10.00)	(10.03)
books (11-25)		-19.99***	-19.38***		-16.85***	-15.72***
		(6.64)	(6.51)		(5.33)	(5.37)
books (26-50)		*Ref.*	*Ref.*			
books (51-100)		20.59***	21.17***		22.00***	22.83***
		(4.79)	(4.72)		(5.26)	(5.28)
books (>100)		44.43***	44.75***		41.22***	41.80***
		(5.73)	(5.65)		(5.82)	(5.84)
parental hisei		1.25***	1.25***		0.97***	0.94***
		(0.15)	(0.15)		(0.17)	(0.17)
migration		-36.83***	-33.75***		-23.22***	-22.94***
		(6.04)	(5.77)		(5.45)	(5.50)
Class/teacher level controls						
% male students			37.36*			19.12
			(21.81)			(20.72)
% language problems			-37.59			-54.09*
			(25.62)			(30.98)
size			0.25			0.14
			(0.69)			(0.66)
teacher further training			-4.55			-6.90
			(4.37)			(4.87)
teacher age			-1.43			0.86
			(2.23)			(2.50)
teacher age2			-0.00			-0.01
			(0.03)			(0.03)
teacher experience			1.04*			0.02
			(0.60)			(0.48)
parttime (Ref = full-time)			7.11			-5.09
			(5.35)			(5.34)
Constant	513.13***	-1,741.86***	-1,716.28***	498.44***	-1,311.82**	-1,189.94**
	(3.63)	(543.67)	(567.79)	(3.36)	(549.75)	(535.08)
Observations	2871	2871	2871	2987	2987	2987
Number of classes	308	308	308	308	308	308
Adjusted R-squared	0.001	0.219	0.228	0.000	0.198	0.204

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply.

Table II: Differences between specifications for boys and girls compared to interaction effects

	Test scores				Grades			
	(3a)	(3b)	(3b) - (3a)	(4)	(8a)	(8b)	(8b) - (8a)	(9)
Reading	-5.69	-11.76*	-6.07	-2.87	0.02	0.03	0.01	0.01
	(7.31)	(6.84)		(6.55)	(0.08)	(0.08)		(0.06)
Math	-4.27	2.46	6.73	8.69	0.06	0.04	-0.02	-0.03
	(5.48)	(6.01)		(6.64)	(0.06)	(0.07)		(0.07)
Science	-8.11	-0.66	7.45	15.52*	-0.03	0.10	0.12	0.15*
	(8.81)	(6.95)		(9.33)	(0.09)	(0.09)		(0.08)

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Figures of specifications are drawn from table 3 (test scores) and table 4 (grades). Differences between by own calculations.

Table III: Likelihood of schools to have a male teacher (Logistic Regressions)

VARIABLES	male German teacher OR (S.E.)	male Math teacher OR (S.E.)	male Science teacher OR (S.E.)
library	0.58	0.90	0.65
	(0.19)	(0.28)	(0.27)
special needs	1.53	0.91	1.10
	(0.91)	(0.51)	(0.76)
shortage of teaching aids	0.71	0.67*	0.79
	(0.18)	(0.19)	(0.22)
disfrac students 0-10%	(=Ref.)	(=Ref.)	(=Ref.)
disfrac students 11-25%	0.70	1.07	0.66
	(0.28)	(0.37)	(0.29)
disfrac students 26-50%	0.99	0.76	1.01
	(0.50)	(0.46)	(0.65)
disfrac students >50%	0.75	1.10	0.77
	(0.46)	(0.67)	(0.63)
rural	(=Ref.)	(=Ref.)	(=Ref.)
suburban	0.84	0.91	0.87
	(0.34)	(0.38)	(0.40)
urban	0.74	0.92	1.04
	(0.30)	(0.37)	(0.58)
Observations	308	308	308
Pseudo R-squared	0.025	0.013	0.012

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are reported in parentheses. Models are averaged over multiply imputed datasets.