

Discussion Paper No. 06-060

**Pupil-Teacher Gender Interaction Effects
on Scholastic Outcomes
in England and the USA**

Andreas Ammermüller and Peter Dolton

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 06-060

**Pupil-Teacher Gender Interaction Effects
on Scholastic Outcomes
in England and the USA**

Andreas Ammermüller and Peter Dolton

Download this ZEW Discussion Paper from our ftp server:

<ftp://ftp.zew.de/pub/zew-docs/dp/dp06060.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Nontechnical summary

The difference between girls and boys academic performance is a major issue on both sides of the Atlantic. The consequences of gender differences in achievement have made the cover story of Newsweek in the form of 'The Boy Crisis' and in the UK there is much consternation each summer when the results of the annual school GCSE exams are published. These figures show how girls are significantly outperforming boys and they are regularly dissected in the media.

To date, the work on this topic has taken an introspective form by examining subjective evaluations of pupils by their teachers or, – at best – nationally determined tests. The objective of this paper is to study the differences in girls and boys achievement using data from different countries which is directly comparable and not subject to the norms and processes of that country's own examination system. One factor which has received increasing attention is the possibility of gender interaction effects, i.e. if boys perform better when taught by male teachers and girls perform better when taught by female teachers. This paper investigates the presence of such 'pupil-teacher gender interactions' on scholastic performance. In this paper we explore this phenomenon at grade 4 and grade 8 in Reading, Maths and Science using TIMSS and PIRLS data for England and the United States.

We find some positive joint pupil-teacher gender interaction effects in the teaching of Maths in England but not the USA in the most recent data for 2003. This would suggest that, all other factors controlled for, there will be a slight advantage to teaching Maths to boys with male teachers and to girls with female teachers in England. An important finding is that there is no gender interaction effect between girls and boys at grade 4 (age 9). This is despite a significant raw difference between girls and boys in Maths and Science in the USA. This suggests that the differences we find between boys and girls and the presence of the gender interaction effect operate at an older age. Perhaps one of the most important implications of these findings is that in order to facilitate any advantage of gender interaction effects we need more male teachers. In both the UK and the US around 75% of teachers are female. Hence it makes sense – if gender interaction effects are present – to recruit more male teachers.

Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA

Andreas Ammermueller*

Centre for European Economic Research (ZEW), Mannheim

Peter Dolton

Royal Holloway College, University of London and CEP London School of Economics

August 2006

Abstract

The difference between girls and boys academic performance is a major issue on both sides of the Atlantic. Do boys and girls fair better with a teacher of their own gender? This paper investigates the presence of such ‘pupil-teacher gender interactions’ on scholastic performance. We use data from PIRLS and TIMSS on Reading, Science and Maths at grade 4 and grade 8 for England and the USA for data from 1995, 1999, 2001 and 2003. We find evidence of gender interaction effects in the form of both positive male interaction effects in Maths scores in the US and Science scores in England at grade 8. Further, using individual fixed effects, Gain score analysis of the difference between Maths and Science scores confirms the presence of Maths gender interaction effects in England (but not the USA) at grade 8 by 2003 when these effects were not present in 1995 or 1999.

Dolton acknowledges financial support from the DfES grant to the Centre for Economics of Education at the London School of Economics and Ammermueller acknowledges financial support from the Fritz Thyssen Foundation under the project “Bildungschancen zwischen Grundschule und Sekundarstufe”.

* Corresponding author: Tel: ++49/621/1235284, email: ammermueller@zew.de

1 Introduction

The difference between girls and boys academic performance is a major issue on both sides of the Atlantic. The consequences of gender differences in achievement have made the cover story of Newsweek in the form of ‘The Boy Crisis’ and concern been expressed by the US First Lady, Laura Bush ‘In almost every case girl’s statistics are better than boys’, ‘I feel like, in the United States, that we’ve sort of shifted our gaze away from boys for the last several decades, and that we’ve neglected boys’. Last year the issue of subject specialization and whether women are less able at Maths and Science was most controversially debated after the remarks of Professor Larry Summers – the then President of Harvard University.¹

In the UK there is much consternation each summer when the results of the annual school GCSE² exams are published. These figures show how girls are significantly outperforming boys³ and they are regularly dissected in the media⁴. They have also been the subject of numerous more serious academic studies (see for example Younger and Warrington (2005), Burgess et al. (2005), Andrews et al. (2006)).

The issue of gender differences in educational attainment is not confined to the US and UK. The same problem arises in many OECD countries as recent OECD figures show. ‘As females have closed the gap and then surpassed males in many aspects of education, there are now many instances in which there is concern about the underachievement of males in certain areas, such as reading. ... Gender differences in student performance need to receive close attention from policy makers if greater gender equity in educational outcomes is to be achieved’.⁵ Recent data from the PISA study shows that girls have a significantly higher performance in reading in 39 out of the 40 countries surveyed (see OECD (2004)).

The importance of gender differences in educational attainment are therefore of great interest. To date, the work on this topic has taken an introspective form by examining subjective evaluations of pupils by their teachers or, – at best – nationally determined tests. The

¹ See the full text of his address at www.president.harvard.edu/speeches/2005/nber.html

² See Annex A where we briefly explain the UK GCSE examination system.

³ Gender differences in these public examinations are published on www.standards.dfes.gov.uk/genderandachievement/understanding/analysis.

⁴ See for example BBC News Online, 22 August 2002, ‘Addressing the Gender gap’ and ‘Exam results reveal gender gulf in schools’ by Amelia Hill in the Observer of May 15, 2005.

⁵ OECD (2003) p10.

objective of this paper is to study the differences in girls and boys achievement using data from different countries which is directly comparable and not subject to the norms and processes of that country's own examination system.

Girls are now outperforming boys in school in most subjects. In the past this was not true for Maths and Science subjects - but increasingly it is true in all subjects. These gender differences have alarmed many educationalists and policy makers. There is a growing literature on the factors which may determine these gender differences in the UK and in the USA (see Entwisle et al. (1994), Burgess et al. (2005), Andrews et al. (2006)). One factor which has received increasing attention is the possibility of gender interaction effects, i.e. if boys perform better when taught by male teachers and girls perform better when taught by female teachers. This paper investigates the presence of such 'pupil-teacher gender interactions' on scholastic performance.⁶

There appears to be two possible explanations abroad for these developments. Firstly, it is suggested that schools and teachers are the source of stereotyping that harm girls particularly in the Sciences and Maths (see Lavy (2004)). Secondly it is suggested that having a teacher of the same sex acts as a role model on pupils (see Bettinger and Long (2005) and Dee (2005)).

One of the important questions which is often raised is whether some of the poor performance of boys is due to the lack of effective male teachers. The argument is that boys need role models and there are so few male teachers now that this is one of the reasons boys are disaffected at school. This concern has been mirrored in the US with the publication of several papers (see Ehrenberg et al. (1995), Bettinger and Long (2005), Dee (2005)) which have reported gender interaction effects between the teacher and the pupil.

In this paper we explore the phenomena of the difference between the performance of girls and boys in school at grade 4 and grade 8 in Reading, Maths and Science using TIMSS and PIRLS data for England and the United States. Specifically we use the TIMSS data from 1995, and 1999 and that which has become most recently available from 2003. This data has the advantage of being internationally standardised data on pupil achievement relating to real knowledge of the pupil as tested in a controlled environment. This is distinct from much of

⁶ Other papers in this literature have also researched the possibility of ethnicity interaction effects e.g. Ehrenberg et al. (1995) but the lack of ethnicity information in our data prevent us investigating this.

the literature on gender interaction effects which focuses on subjective evaluations of pupil performance by their own teachers. This means we are interested in real differences in attainment between girls and boys rather than perceived differences as judged by their teachers. We are therefore not interested in the possibility of stereotyping but rather the real academic or scholastic effect of having a teacher the same gender as the pupil. The form of the data means we may make effective comparisons between different countries. In this research we focus only on the comparison of the USA and England as it is in these countries that most of the evidence relating to gender comparative performance is available.

One important limitation to this study which must be borne in mind is that we do not have knowledge of the gender of the pupil's teacher in previous years. Hence what we seek to identify is the effect of the gender of the pupil's teacher in the current year that they sit these international tests. In some respects then we are asking the data to identify what we may expect to be a small effect - namely the effect of the gender interaction on current performance. Hence should we find such effects our estimates are likely to be severely understated due to the fact that we do not observe the gender of the past teachers of each pupil. This means that should we find such gender interaction effects in this context we should appreciate that the real effect of this phenomena are likely to be much larger than we find in our data.

2 The Literature

Literature on Gender differences in pupil academic achievement

There is a large literature on the difference between the academic achievement of boys and girls. Historically there was always a gap favouring boys in Maths and the Sciences and a gap favouring girls in reading, English and the arts and languages. This has been well documented in many places (e.g. Entwisle et al. (1994)). In recent times this pattern has been changing with girls improving in the Maths and Sciences and even outperforming boys in many countries.

In the UK there is widespread evidence of girls outperforming boys at most levels in school and specifically there is now a wide gap between boys and girls by the age of 14-16 in the GCSE public examination results. The divergence of performance between boys and girls clearly began in 1987 when the public examinations system changed from O levels to GCSE. This pattern is described in Dolton et al. (1999) and been examined in detail by Burgess et al

(2005), Machin and McNally (2005) and Andrews et al. (2006). The essential detail is that the gap between girls and boys (in favour of girls) has always been present in English and is now getting wider and that girls have improved in Maths and are now outperforming boys (see Machin and McNally (2005) for details). The problem with the UK literature is that it is difficult to assess the extent to which the difference between boys and girls is due to the structure of the examination system *per se*. It is clear from Dolton et al. (1999) and Andrews et al. (2006) that the changes in the learning and assessment methods inherent in the system of school exams has had a lot to do with the development of this gap. Ideally we would like to know how the abilities and performance of boys and girls have moved independently of the changes in the exam system. This is not possible, using data on Key Stage 2, 3 or 4⁷ or GCSE results, as performance on these tests is now directly associated with the form of these tests. Unfortunately it is impossible to distinguish how much of the difference between boys and girls is due to genuine patterns of change in the improving performance of girls rather than the form of the examination system. Although the evidence suggests that different methods of assessment are not gender neutral (see Powney (1996)), and more specifically that a higher weighting of course work elements improves the relative performance of girls (see Stobart et al. (1992)). We would also like to be able to compare the performance of boys and girls in the UK with other countries. The PIRLS and TIMSS data afford us the opportunity to do just this and to exclude any impact of teacher assessment.

In the US the debate over the differences in the educational attainment of boys and girls has studied perennially and been consistently in the news and popularist press. Ever since Macoby and Jacklin (1974) (and before) the differences between girls and boys on verbal, mathematical and scientific tests has been documented. The broad consensus was that boys outperform girls on Math and Science but that girls do better at verbal and linguistic tests. This finding holds up in a most recent study with excellent data which is able to control for a diverse set of student characteristics and a comprehensive set of teacher characteristics (see Clotfelter et al (2006))⁸. Some recent evidence (Linn and Hyde (1989) and Entwisle et al. (1994)) has been that these gender differences in attainment have been declining. The most recent reported media ‘crisis’ of boys has been the result of the latest psychological studies which concentrate on the physiological differences between boys and girls brains in childhood and adolescence (see Sax (2005)) and descriptions of the rise in drug use to control wayward

⁷ See Annex A for an explanation of the UK Public Examination System.

⁸ In this study they do not control for pupil-teacher gender interaction although the simple effect of teacher gender seems to be negligible when a reasonable array of controls are included.

behaviour in the classroom by young males (see Diller (1998)), the accounts of how girls have received the most help from government funding (see Sommers (2000)) and a call for more single sex educational provision (see US Department of Education (2005)). This has met with the outcry from Newsweek (2006), the knee jerk reaction of the White House, and a more balanced ensuing discussion in the media (see Hulbert (2004; 2006) and Rivers and Barnett (2006)).

Notwithstanding the raw difference in the performance of boys and girls in school we are more specifically interested in this paper in the possible identification of the effect of pupil-teacher gender interactions on academic performance. One of the suggestions of the Newsweek article was: 'One of the most reliable predictors of whether a boy will succeed or fail in high school rest on a single question: does he have a man in his life to look up to?' This remark was of course also a reference to the presence or absence of fathers in modern families – but the suggestion remains that male teachers may make a difference to boys' educational progress. There is a smaller, more focussed literature, on this topic.

Literature on Gender Interaction Effects

Ehrenberg et al. (1995) use the National Educational Longitudinal Study (NELS) of 1988 in the US to study the relationship between gender and ethnicity pupil teacher interactions of pupils performance and teacher evaluations of the pupils. Essentially they found no evidence to support the hypothesis that same sex or same race teachers improve the test scores of their pupils. Their analysis was in terms of the difference in 8th grade performance by gender or ethnicity in History, Reading, Math or Science. Their analysis was in terms of 'gain scores', where this was measured in terms of the difference in performance at the 8th grade and the 10th grade in each subject. They then compared this to the teacher's subjective evaluation of the pupil at the 10th grade in each subject. In these evaluations there is evidence that the gender and race of the teacher do have significant interaction effects with the gender and race of the pupil. They found that test scores of white female students in Maths and Science did not increase more rapidly when the teacher was a white woman than when the teacher was a white man. However they did find that white female teachers did evaluate their pupils more highly.

The evidence of Ehrenberg et al. (1995) does not sit comfortably with that of Dee (2005) who also uses the NELS 1988 data to investigate gender and race interaction effects. However Dee

(2005) does not assess academic gain scores (whether measured subjectively or not) but rather focuses on differences in teacher perceptions of student behaviour.⁹ His identification strategy comes from having two separate subject assessments from different teachers on each student. He concludes that racial, ethnic and gender dynamics between students and teachers have consistently large effects on teacher perceptions of student behaviour. The teachers in the survey were asked to judge whether individual students in the sample were seen as frequently disruptive, consistently inattentive or rarely handed in homework. He uses these dichotomous assessments as binary dependent variables in a fixed effects logit model where he regresses them on whether the student and teacher are of the same gender and race or not. He consistently finds that students are more likely to be judged by the teacher to be disruptive, inattentive or rarely hand in homework if they are of another gender or race than the teacher. He finds these effects are most significant for low SES students living in the South. The problem with this identification strategy in this study is that it relies on teacher assessments being independent and there being no differences in the natural abilities or interests of the students across certain subjects that are not accounted for by other than the teacher's race or gender.

Lavy (2004) investigates whether gender stereotypes affect teachers' scores for pupils on matriculation examinations. Using data from both blind and non-blind test scores in several subjects he finds – contrary to expectations – that male students face discrimination in every subject. He also finds that the size of this bias is very sensitive to teacher's characteristics which suggests that the bias against male students is the result of teachers' and not the students' behaviour or characteristics.

Bettinger and Long (2005) address the question of whether the gender of the instructor for early college courses influences subsequent choice of course subjects. They suggest that in subjects like maths, geology, sociology and journalism there is evidence of support for the role model theory that having a same sex tutor induces a person to be more likely to choose the subject. Although their focus of choice of subject rather than academic performance is not quite the same issue it is likely that subject choice correlates well with prior performance and this may be at the heart of the gender differences which were commented on by Larry Summers.

⁹ Dee (2004) has another paper that does look at academic achievement using the Tennessee project STAR data but concentrates on the use of this data to identify race interaction effects.

Holmlund and Sund (2005) use data on upper secondary school pupils in Sweden to explore the gender difference in performance. They find that the gender performance differential is larger in subjects where the share of female teachers is higher and explore the possibility that there is a form of endogeneity between the gender balance of entry into teaching and the outcomes of the pupils by gender. Their estimation strategy is to compare the OLS estimates of scores in English, Swedish and Maths (for those in two programmes - Natural science and Social science) with the 'within-student' estimates of achievement gain over three years for the same students. This strategy allows for the effect of previous teachers in earlier years to impact on current scores. They conclude that there is no support for the hypothesis that a same-sex teacher improves student outcomes.

There are some clear limitations to the Holmlund and Sund analysis - firstly they only have data relating to 16-18 years old pupils in Sweden - if we expect to have teacher pupil interaction effects we would expect them to operate at an earlier age than this. Secondly, their grade outcome data only takes 4 values 0, 10, 15 and 20 which correspond to teachers' qualitative subjective assessments of the pupils. This is not ideal if we wish to identify same gender interaction effects as it automatically means that we do not know if the difference between the performance of girls and boys is genuinely due to a difference in their performance or due to the way in which they have been assessed by their teachers. In addition the use of qualitative assessments rather than objective exam or test scores means that there is scope for substantial measurement error bias. If what we are really interested in is the pupil's actual performance and test scores there is limited value to the subjective qualitative assessments of the teachers.

It can be seen from this short review of the empirical literature that most of the existing studies on gender interaction effects relate to examination systems which are internal to a country and rely on data relating to subjective assessments by the teachers who were teaching the pupils. In addition there is no possibility of a genuine inter-country comparison of the gender effects in different subjects across time. The results we report in this paper attempt to meet these criticisms by providing evidence from objectively set exams which are internationally comparable.

3 Econometric Identification and Estimation

Consider the model used by Holmlund and Sund (2005):

$$y_{ijk} = \alpha + \beta_1 W_i + \beta_2 T_i + \beta_3 I_{ijk} + \gamma' X_{ijk} + \eta_j + \lambda_k + \varepsilon_{ijk} \quad (1)$$

where y_{ijk} is the test score for student i in school j in subject k

W_i is a dummy variable taking the value 1 if the pupil is female.

T_i is a dummy variable taking the value 1 if the teacher of the i 'th pupil is female

I_{ijk} is an interaction term between the female pupil and the female teacher

X_{ijk} is a vector of controls

η_j captures school effects

λ_k is a set of indicator variables for each subject

ε_{ijk} is a normally distributed iid error term.

The problem with the specification in (1) is that it does not allow us to identify the interaction effect of male teachers with male pupils. It also assumes that the other coefficients in the model, γ , are common between males and females. This is an unnecessary restriction.

In their OLS model Holmlund and Sund are concerned that I_{ijk} is not independent of ε_{ijk} .

This problem would arise if teachers of a specific gender are consistently assigned pupils of a higher potential performance or if there is some other way in which the assignment of teachers (by gender) is non-random across pupils of differing abilities. It is for this reason that Holmlund and Sund estimate a within student model of a fixed effects form of equation (1) between two time periods. There are also problems with this specification since it still assumes that the effects of the time varying regressors are constant across boys and girls. It also assumes that any individual specific unobserved heterogeneity relating to pupil performance across time is constant. This would seem unreasonable in a 2 year time frame used by Holmlund and Sund.

To address these issues in our data we will run separate estimations for boys and girls - as in equation (2). If we can interpret the assignment of girls and boys to classes with male or female teachers as a random treatment then it will be possible to interpret the β^b, β^g as the treatment effects of boys having male teachers and girls having female teachers.

$$\begin{aligned}
y_{ijk}^b &= \alpha^b + \beta^b I_{ijk}^b + \gamma' X_{ijk}^b + \eta_j^b + \lambda_k^b + \varepsilon_{ijk}^b \\
y_{ijk}^g &= \alpha^g + \beta^g I_{ijk}^g + \gamma' X_{ijk}^g + \eta_j^g + \lambda_k^g + \varepsilon_{ijk}^g
\end{aligned}
\tag{2}$$

We will report these estimated interaction effects on the assumption that they are treatment effects which result from random assignment of classes. An identification problem would arise if the assignment of teachers by gender was not random to either high or low ability classes. Indeed, our identification would also be problematic if there were any other teacher characteristics which were unobserved which may relate to pupil performance in a way which was not gender neutral. This concern clearly drives the Holmlund and Sund (2005) study. A second possible identification problem arises with the interpretation of the interaction effects if we cannot unambiguously attribute the effect of estimated interactions as being only the effect of having a teacher of the same sex as clearly any academic achievement is the result of the current year's learning environment as well as the previous year's environment (cf. Todd and Wolpin, 2003). Hence any gender interaction effect this year could be swamped by previous year's effects. To recap- in order to interpret our findings as identifying the gender interaction effect we must make three identification assumptions:

1. That the assignment of gender of the teacher is independent of the gender of the pupil.
2. There are no unobserved teacher characteristics which may affect performance of girls and boys differently.
3. That the effect of the teacher pupil interaction is a marginal effect of the gender of the pupil in the current academic year on the academic score of the pupil in that year and not a product of previous year's teacher's efforts.

If we do not accept that assignment of the gender of a class teacher is random then the identification of the presence of gender interaction effects is not so straightforward. One possible way forward is if we have repeated measures of the academic performance of the students. In the case of Holmlund and Sund (2005) they use repeated measures of the student two years apart. In the case of Dee (2005) he uses two measures for the same student in different subjects. We adopt the same identification strategy here. Notice however that we will still need to assume that the interaction effect is all associated with this year's teachers rather than the teachers of previous years, i.e. the identification assumption 2 above. With this in mind we can write the Gain score of achievement in Maths over the achievement in Science for each individual as:

$$\begin{aligned}\Delta y_{ij}^b &= (y_{ijm}^b - y_{ijs}^b) = \alpha^b + \beta_m^b I_{ijm}^b - \beta_s^b I_{ijs}^b + \gamma' \Delta X_{ij}^b + \varepsilon_{ij}^b \\ \Delta y_{ij}^g &= (y_{ijm}^g - y_{ijs}^g) = \alpha^g + \beta_m^g I_{ijm}^g - \beta_s^g I_{ijs}^g + \gamma' \Delta X_{ij}^g + \varepsilon_{ij}^g\end{aligned}\tag{3}$$

Notice that this gain score estimation strategy will get rid of all the covariates which are constant and at the same time eliminate any individual specific unobserved heterogeneity. Only teacher characteristics differ between the two subjects and are still included as covariates.

A logical alternative estimation of interest is to pool boys and girls in the equation above and report ‘joint’ interaction coefficients which simply measure the interaction effect of the same gender as the pupil irrespective of whether this refers to girls and female teachers or boys and male teachers. We report these coefficients as well in our results.

4 The Data

The PIRLS data

Thirty-five 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 and tested fourth grade students (nine- and ten-year-olds) in reading literacy. In the data, extensive information on home and school environments is available through student, parent, teacher and school questionnaires. With 150,000 students tested, PIRLS 2001 is the first in a planned 5-year cycle of international trend studies in reading literacy (Mullis et al. (2003)).

The data are clustered due to the stratified sampling design of the study. The schools that participated have been chosen first, before a sample of classes from the targeted grade was drawn. Therefore, the schools are the primary sampling units and not the classes or students. Student performance is measured by test scores in reading literacy, which is the most important basic competency needed to acquire further skills and knowledge and to successfully participate in social life (Mullis et al. (2003)). The test scores are plausible values that are drawn from an estimated proficiency distribution. Plausible values are imputed scores based on the students’ answers to the test items (cf. Mislevy (1991)). The scores have then been standardized, to an international mean of 500 and a standard deviation of 100, which facilitates the comparison across countries.

The TIMSS data

The Trends in International Mathematics and Science Study (TIMSS) was conducted by the IEA in about 50 countries worldwide in 1995, 1999 and 2003. This publicly available data combines student-level performance data in mathematics and science and extensive background information. Students in both primary and middle secondary education were tested at grade 4 (around 9 years old) and grade 8 (around 13 years old). In 1995, students in the adjacent grades eight and nine were tested in England and in grades seven and eight in the US. For 1999, data is available on students in grade nine in England and grade eight in the US. In 2003, students were tested in grades four and eight in both countries. Students in grades seven, eight and nine are referred to as grade eight students. All available data for an individual student were merged together with his sampling weight (Martin et al. (2004)).¹⁰

The PIRLS and TIMSS data are based on studies that have been conducted by the same organization (IEA), use the same sampling design and similar questionnaires. In addition the same definitions of variables are used and the test scores are computed likewise. Table A1 presents selected statistics.

Variables

Student background is described by five variables besides student's gender. Student's age is measured in years up to the first decimal point. Since we compare students within a particular grade and not students of the same age group, it is important to control for differences in age that indicate both the age at school enrolment and a possible grade repetition. When students are tested in two adjacent grades in 1995, controls for the grade level are added. Student's migration background is described by two dummy variables. One indicates whether at least one parent was born abroad while the other shows whether students speak a language other than the test language at home. The variables on migration background hence include both information on the origin of a student's family and the degree of integration. Information on the country of origin is not available. The number of books at home consists of five categories: ten and fewer books (reference category), 11-25 books, 26-100 books, 101-200 books and more than 200 books at home. The number of books at home represents the social and cultural background of students and is preferred to parents' education because it is the single most influential determinant of student test scores and has fewer missing values. A further variable comprises the number of persons besides the child that live in the household, which is an indicator of the family structure.

¹⁰ For more information on PIRLS and TIMSS, see the TIMSS internet homepage <http://timss.bc.edu>.

Several variables describe teacher characteristics. Besides teacher's sex, teacher's experience and experience squared are included as explanatory variables. (Moreover, information is available on the highest level of completed formal education of teachers, ISCED 5a first degree or second degree. It is also indicated whether teachers have a major in either mathematics or science.)

The size of the community where the school is located is included to control for social and cultural differences between rural and urban areas. The reference category is town (3,001 – 100,000 inhabitants); the two further categories are city (more than 100,000 inhabitants) and rural (less than 3,001 inhabitants).

There are many caveats to the use of this data for the analysis of gender interaction effects and so we must proceed cautiously in our interpretations of the results. Specifically it will be hard to make a comparison on scores in TIMSS with performance in national examinations. It should be remembered that TIMSS is based mostly on multiple choice questions and hence are structurally very different from exams like GCSE's in the UK which are usually a combination of marks on course work, and examinations which may be essay type questions or exams which require written responses in structured paragraphs rather than simply picking an alternative from multiple possibilities.

We choose a parsimonious specification for our production function in order not to drop too many observations due to missing values. The variable class size is not included in the specifications we present here due to a sizeable fraction of missing values in the US. The results are very similar for alternative specifications including more control variables like teacher's formal education level, though.

5 Estimation Results

Prior to any regression analysis it is in order to examine the raw unconditional scores of the boys and girls on their Maths, Science, and Reading tests at grade 4 and 8 in PIRLS (2001) and TIMSS (1995, 1999, and 2003). Tables 1 and 2 present these raw statistics for the whole sample.¹¹ We see from these tables that boys outperform girls in Maths and Science at grade 4 in the USA but the girls outperform boys in Reading in both England and the USA. By

¹¹ Tables A2 and A3 present the same statistics for the most restrictive sample used in the estimation in Table 3.

grade 8 boys outperform girls in both countries in Science and Maths in 1995 and 1999 but that it is only in the USA that boys outperform girls significantly in Maths by 2003. Looking at the summary statistics for GAIN scores in Table A2 we see that both boys and girls perform better in Science than Math but that this difference is significantly higher for boys than it is for girls in all years of the TIMMS data. Further examination of Table 2 shows that the boys' advantage over girls in Maths and Science at grade 8 is mostly smaller in 2003 than it was in 1995 or 1999. Indeed in England by 2003 boys have no significant advantage over girls in Maths. Clearly girls on both sides of the Atlantic have been bridging the gap that used to exist. It should be stressed that this table is only of the raw differences unconditional on any characteristics of the individual (or their home) or the school or the class. However the contrast with the position at age 16 GCSE in the UK should be noted. Specifically we should reflect on how girls have such a large advantage by this age in these exams when it is clear that the gap is reversed at age 13 in TIMSS. We will return to this question.

Further light can be shed on the differences between the achievement pattern differences between boys and girls by consideration of the whole distribution of scores in the different subjects at the different grades in England, and the USA. The kernel plots of these distributions are graphed in the Data Appendix in Figures A1 to A3. The pattern which seems to emerge from these plots is that boys have a wider variance of performance in Maths and Science in both countries with more mass in the tails of the distribution than girls at grade 4. In contrast girls have a distribution of reading scores which is uniformly to the right of boys on the PIRLS reading test. Looking at the same kernel plots for both Maths and Science in both countries for the grade 8 scores it would seem that the boys' distribution of scores has a lower mass around the median with a shift to the right compared to the girls' distributions.

The conditional regression analysis tabulated in Tables A4 to A8 report on a range of effects of the home environment and the school. Most explicitly we see that the scores are directly affected by the parents language used at home and how many books there are in the home environment. There are also adverse household size and inner city effects¹². This is not the focus of our analysis here, however. Rather we wish to concentrate on the interaction effects reported in Table 3 which detail the OLS results on the β^b, β^g coefficients from equation (2) and the fixed effects estimates of the Gain score analysis $\beta_m^b, \beta_m^g, \beta_s^b, \beta_s^g$ of equation (3) which are tabulated in Table 4.

¹² These kind of results are well known from the results of Woessmann (2004) and others.

Table 3 reveals interesting interaction effects are present in 2003 from the TIMMS data. The OLS model shows there is a significant positive effect for boys in the USA in Maths at grade 8 and a significant positive effect for boys for Science at grade 8 in the UK in 2003. These results contrast with the TIMMS data from 1995 and 1999. In 1999 there is a positive interaction effect for females in Science in England. Hence the finding of a strong Maths effects for boys in the USA by 2003 is rather surprising. The size of the effect corresponds to around a quarter of a standard deviation which may be the equivalent of shifting an individual persons position in the score distribution up by around 7 or 8 percentage points. This is a rather startling effect and the explanation for it is unclear and worthy of further investigation.

Perhaps most interesting of all are the interaction effects in the Gain score analysis using fixed effects¹³. It can only be conducted for TIMSS in grade 8, where students are tested in two subjects and have different teachers. Here we see a positive gender interaction for girls in Maths in England in 2003. For 1995, we find a negative interaction effect for boys in Maths in the US. For 1999, there is a positive interaction effect for boys in Sciences in England. The size of these effects are understandably much smaller as they represent the Gain score in Maths over the score in Science for the same pupil for having a male Maths teacher if you are a boy. One would expect ability in Maths and Science to be highly correlated and hence to find any significant impact on the difference in the scores in these two subjects is a rather startling result. In addition to a boys gender interaction effect for Sciences – there is also one for girls in England in Maths in 2003. This suggests therefore that girls fair better in Maths with a female teacher. Table 4 also shows us that this result is a fairly recent phenomenon as there is no Maths effect result for earlier years.

The most consistent findings reported in Table 4 come when we ran a restricted version of our model in equation (3) restricting the gender interaction effects to be the same for boys and girls. We report this as a third pair of rows in each section of Table 4. We call these 'joint interaction effects'. The results suggest that there is a Maths strong positive interaction effect in England but not in the USA in 2003. This effect was also present in the USA in 1999 but not in England. These results can be contrasted with those for Science which were negative in

¹³ The distribution of these Gain scores is graphed in Figure A4. This shows that scores in Science at the individual level are typically higher than those for the same person in Maths. This is more true in the USA than England.

the USA in 1999 and disappeared by 2003. The Science interaction effect is much less evident in England with no significance by 2003.

One possible mechanism which could partially be driving these results is that girls and boys in single sex schools have better schools and since they are more usually taught by teachers of the same gender we could find that the evidence for the presence of gender interaction effects is weaker when we exclude the pupils who are at single sex schools. We tabulate these results in Table 5. Indeed this is what we find – when we estimate the GAIN score model for only those in co-educational schools. This implies that part of the gender interaction effect may be due to the greater effectiveness of single sex schools.

Taken together the findings of the GAIN score analysis do suggest there is evidence that there are positive academic achievement returns to a child having a teacher of the same gender. If we do not condition out for unobserved heterogeneity with our fixed effects analysis then there is some evidence of interaction effects particularly for boys having male science teachers in England. There may also be a return to girls having female maths teachers. This effect is to be found in the most recent data collected in 2003. However they are not to be found in earlier data from 1995. Indeed there is even a suggestion that gender interaction effects could have been negative in the US in 1995. It is unclear what has prompted this dramatic change.

6 Policy Context and Implications

One of the important implications of these findings is that potentially too much is being made of the educational underachievement of boys particularly in the UK and the US. The large differences in the performance of boys and girls at GCSE in the UK and the differences in boys and girls performance in school SAT or state wide tests in the US are not mirrored in the TIMSS tests where boys have a slight raw advantage over girls. This would lead to the suggestion that the existing internal country specific exam systems favour girls and the way in which they learn and study. This is good news in the sense that our policy agenda should not be so concerned with how to adapt educational learning practices to help boys learn. Rather we should be interested in considering the way in which the present structure of tests and exams favours girls and how to make these exams gender neutral. To counter this conclusion one should appreciate that TIMSS and GCSE are measuring very different things. Specifically TIMSS takes place at age 13 whereas GCSE takes place at age 16. Over these

crucial years of adolescence boys and girls mature at very different rates. Hence we may be picking up the fact that there is no 'boy problem' at age 13, but that one appears by age 16. An alternative explanation is that boys can do multiple choice exams but girls are much better at assessed course work and essay questions in examination conditions. Evidence is also mounting on the amount of time boys spend playing computer games, truanting from school and not doing their homework.

An important finding is that there is no gender interaction effect between girls and boys at grade 4 (age 9). This is despite a significant raw difference between girls and boys in Maths and Science in the USA. This difference took the form of boys outperforming girls by up to a 1/10th of a standard deviation. However, when family and educational factors are controlled for these differences between boys and girls are not due to gender interaction effects. This suggests that the differences we find between boys and girls and the presence of the gender interaction effect operate at an older age. This would imply that there is no premium to a teacher of the same gender when pupils are of primary age but there is a more pronounced effect when the child is of secondary age.

The main focus of this paper was to test to see if there are any gender interaction effects in the performance of boys and girls in Science, Maths and Reading. We found that there were some positive joint pupil-teacher gender interaction effects in the teaching of Maths in England but not the USA in the most recent data for 2003. This would suggest that, all other factors controlled for, there will be a slight advantage to teaching Maths to boys with male teachers and to girls with female teachers in England. A logical argument from these findings is that they may support single sex education – provided that boys taught separately are taught by male teachers and girls by female teachers.¹⁴ This would not be a correct conclusion to draw from this work as we have not been able to control for the effects of being taught in a single sex class. We have only focused on being taught by a teacher of the same sex – irrespective of what the gender balance of pupils in the class is. Much more research with more detailed data would be necessary to form this conclusion.

Perhaps one of the most important implications of these findings is that in order to facilitate any advantage of gender interaction effects we need more male teachers. In both the UK and

¹⁴ There is a considerable lobby for single sex education in the US - see Rhulman (1996) and US Education Department (2005). Separate evidence from studies of single sex schools suggest that single sex girls schools outperforming boys schools is an additional argument in favour of more single sex schools. See Woodhead (2002) p20.

the US around 75% of teachers are female. Hence it makes sense – if gender interaction effects are present – to recruit more male teachers. This is not as straightforward as it sounds and has definite implications for the wage structure which would be necessary to induce more men into the profession.

One further important caveat to our work is that we were hunting for the potentially very small effect of the interaction of the gender of the teacher with the gender of the pupil in the current school year. If we knew of data sets in which the gender of the teacher was known (and performance of the pupil measured) in many previous years then we could attempt a more sophisticated panel analysis. An alternative identification strategy is to examine countries in which it is more common for pupils to have the same teacher for several years running to see if the gender interaction effects are stronger than those countries in which a pupil routinely changes their teacher each year. This is a topic for further analysis.

References

- Andrews, M., Bradley, S., Stott, D., and Taylor, J. (2006)** 'The Evolution and Determinants of the Educational Gender Gap in England', University of Lancaster mimeo.
- Bettinger, E., and Long, B. (2005)** 'Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students', *American Economic Review*, vol.95(2), pp.152-57.
- Burgess, S., McConnell, B., Propper, C., and Wilson, D. (2005)** 'Girls Rock, Boys Roll: An Analysis of the Age 14-16 Gender Gap in English Schools', University of Bristol, CMPO Working Paper no 03.084.
- Clotfelter, C., Ladd, H. and Vigdor, J. (2006)** '*Teacher-Student Matching and the Assessment of Teacher Effectiveness*', NBER working paper 11936.
- Dee, T. (2004)** 'Teachers, race and Student Achievement in a Randomized Experiment', *Review of Economics and Statistics*, vol.86(1), pp.195-210.
- Dee, T. (2005)** 'A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?', *American Economic Review*, vol.95(2), pp.158-65.
- Diller, L. (1998)** '*Running on Ritalin: A Physician Reflects on Children, Society and the Performance in a Pill*', Bantam, USA.
- Dolton, P., Makepeace, G., Hutton, S. and Audas, R. (1999)** '*Making the Grade: Education, the Labour Market and Young People*', Joseph Rowntree Foundation, York.
- Ehrenberg, R., Goldhaber, D., and Brewer, D. (1995)** 'Do Teacher's Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988', *Industrial and Labor Relations Review*, vol.48(3), pp.547-61.
- Entwisle, D., Alexander, K. and Olson, L. (1994)** 'The Gender Gap in Math: Its Possible Origins in Neighborhood Effects', *American Sociological Review*, vol.59(6), pp.822-38.
- Holmlund, H. and Sund, K. (2005)** 'Is the Gender Gap in School Performance Affected by the Sex of the Teacher?', Swedish Institute for Social Research, Stockholm University, Working paper 5.
- Hulbert, A. (2004)** '*Raising America: Experts, Parents and a Century of Advice about Children*', Vintage Books, USA.
- Hulbert, A. (2006)** 'Will Boys be Boys?: Why the Gender Lens may not shed Light on the Latest Educational Crisis', Slate, <http://www.slate.com/id/2135243>.
- Lavy, V. (2004)** 'Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment', Hebrew University of Jerusalem, mimeo.
- Linn, M. and Hyde, J. (1989)** 'Gender, Maths and Science', *Educational Researcher*, vol.18(8), pp.17-27.
- Machin, S. and McNally, S. (2005)** 'Gender and Student Achievement in English Schools', *Oxford Review of Economic Policy*, vol.21(3), pp.357-372.
- Macoby, E. and Jacklin, C. (1974)** '*The Psychology of Sex Differences*', The Stanford Press.
- Martin, M., Mullis, I. and Chrostowski, S. (Eds) (2004)** '*TIMSS 2003 Technical Report*', Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mislevy, R.J. (1991)**, 'Randomization-Based Inference About Latent Variables From Complex Samples', *Psychometrika* vol.56, pp.177-196.
- Mullis, I., Martin, M., Gonzalez, E. and Kennedy, A. (2003)**, '*PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*', Chestnut Hill, Boston College.
- Newsweek (2006)** 'The Boy Crisis', cover story, Jan 22.

- OECD (2003)** *'Education at a Glance 2004'*, OECD, Paris.
- OECD (2004)** *'Learning for Tomorrow's World: First Results from PISA 2003'*, OECD, Paris.
- Powney, J. (1996)** 'Gender and Attainment - A Review', Edinburgh, Scottish Council for Research in Education.
- Rivers, C. and Barnett, R. (2006)** 'The Myth of 'The Boy Crisis'', Washington Post, Sunday 9th April.
- Ruhlman, M. (1996)** *'Boys Themselves: A Return to Single Sex Education'*, Henry Holt, New York.
- Sax, L. (2005)** *'Why Gender Matters: What Parents and Teachers really Need to Know about the Emerging Science of Sex Differences'*, Doubleday.
- Sommers, C. (2000)** *'The War Against Boys How Misguided Feminism is Harming our Young Men'*, Simon and Schuster, New York.
- Stobart, G., Elwood, J. and Quinlan, M. (1992)** 'Gender Bias in Examination: How Equal are the Opportunities?', *British Educational Research Journal*, vol.18(3), pp.261-76.
- Todd, P. and Wolpin, K. (2003)** 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *The Economic Journal*, vol.113, F3-33.
- US Education Department (2005)** *'Single-Sex versus Coeducation Schooling: A Systematic Review'*, Washington. (see www.eg.gov/rschstat/eval/other/single-sex/index.html)
- Woessmann, L. (2004)** 'How Equal are Educational Opportunities? Family Background and Student Achievement in Europe and the United States', IZA Discussion paper 1284.
- Woodhead, C. (2002)** *'Class War: The State of British Education'*, Little Brown, London.
- Younger, M. and Warrington, M. (2005)** *'Raising Boys' Achievement in Secondary Schools: Issues, Dilemmas and Opportunities'*, Maidenhead: Open University Press.

Table 1: Average scores by country and gender, PIRLS 2001

	Average score (Std. dev.)		Difference (Std. err.)
	Female	Male	
England	562.59 (79.98)	540.29 (83.99)	-20.30^{***} (2.92)
USA	553.47 (76.80)	536.99 (83.88)	-16.48^{***} (2.62)

Average reading scores for female and male students and difference between male and female students for all observed students. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

Table 2: Average scores by subject, country and gender for TIMSS 1995, 1999 and 2003

	MATH			SCIENCE		
	Average score (Std. dev.)		Difference (Std. err.)	Average score (Std. dev.)		Difference (Std. err.)
	Female	Male		Female	Male	
TIMSS 1995						
<i>Grade 8</i>						
England	478.93 (84.84)	488.87 (86.29)	9.94^{***} (2.89)	505.65 (86.78)	526.19 (91.45)	20.54^{***} (3.01)
USA	471.63 (1.13)	477.31 (89.10)	5.68^{***} (1.66)	482.75 (95.20)	499.28 (101.15)	16.53^{***} (1.89)
TIMSS 1999						
<i>Grade 8</i>						
England	491.03 (76.75)	510.45 (81.84)	19.42^{***} (2.98)	526.27 (82.61)	558.75 (86.23)	32.48^{***} (3.17)
USA	493.90 (80.20)	500.57 (87.24)	6.66^{***} (1.79)	498.25 (85.17)	517.18 (95.78)	18.92^{***} (1.93)
TIMSS 2003						
<i>Grade 4</i>						
England	529.79 (80.13)	532.85 (86.93)	3.06 (2.83)	542.98 (75.07)	539.41 (81.83)	-3.57 (2.66)
USA	508.54 (71.84)	515.63 (75.12)	7.08^{***} (1.48)	525.93 (74.93)	530.51 (80.10)	4.58^{***} (1.57)
<i>Grade 8</i>						
England	508.74 (76.00)	508.81 (79.90)	.08 (3.01)	547.48 (71.10)	559.39 (78.49)	11.91^{***} (2.89)
USA	501.18 (75.28)	507.78 (80.04)	6.59^{***} (1.64)	517.84 (75.70)	535.44 (79.55)	17.60^{***} (1.65)

Average scores for female and male students and difference between male and female students for all observed students. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

Table 3: Gender interaction effect estimates

	England			USA		
	(1)	(2)	(3)	(1)	(2)	(3)
PIRLS 2001						
<i>Grade 4</i>						
Reading						
Female interaction effect	2.14 (12.80)	4.48 (10.17)	4.85 (10.57)	-.61 (8.60)	-2.14 (7.37)	-0.80 (9.08)
Male interaction effect	-7.10 (8.98)	-7.54 (7.44)	-6.71 (7.79)	-18.13 (15.80)	-16.41 (12.97)	-19.15 (12.60)
TIMSS 1995						
<i>Grade 8</i>						
Math						
Female interaction effect	17.00 (10.47)	11.65 (8.67)	15.97* (8.37)	1.11 (7.72)	1.39 (5.69)	2.42 (6.13)
Male interaction effect	11.36 (12.76)	13.28 (10.46)	11.89 (11.34)	-5.33 (7.55)	-6.76 (5.52)	-6.06 (6.20)
Science						
Female interaction effect	5.79 (9.53)	5.71 (7.55)	10.75 (7.89)	4.59 (8.29)	4.58 (6.05)	6.33 (6.28)
Male interaction effect	11.48 (13.69)	9.97 (10.81)	4.15 (11.62)	-9.35 (8.18)	-9.70 (6.79)	-10.70 (6.77)
TIMSS 1999						
<i>Grade 8</i>						
Math						
Female interaction effect	11.31 (14.97)	10.94 (12.39)	12.80 (12.10)	-1.91 (7.49)	-3.57 (6.20)	-4.13 (6.16)
Male interaction effect	20.97 (13.35)	14.70 (10.61)	10.06 (10.24)	1.72 (9.70)	4.35 (8.02)	3.92 (8.37)
Science						
Female interaction effect	28.40** (11.55)	16.03* (8.99)	17.03** (8.15)	-0.81 (7.77)	-0.22 (5.96)	-1.76 (6.33)
Male interaction effect	7.74 (12.79)	3.59 (8.94)	1.00 (8.42)	7.07 (8.59)	6.17 (6.74)	6.35 (6.83)
TIMSS 2003						
<i>Grade 4</i>						
Math						
Female interaction effect	-14.61 (11.28)	-6.79 (8.84)	-0.50 (9.19)	-6.79 (8.39)	-7.64 (6.45)	-8.60 (6.19)
Male interaction effect	20.43 (12.81)	12.76 (10.53)	8.65 (10.50)	2.56 (10.14)	1.75 (7.49)	2.57 (7.08)
Science						
Female interaction effect	-18.10* (10.42)	-10.51 (7.44)	-8.81 (8.11)	-4.03 (8.52)	-6.21 (6.48)	-7.92 (6.37)
Male interaction effect	11.53 (10.16)	4.38 (7.99)	1.99 (8.95)	-1.98 (10.82)	-2.21 (7.75)	-0.23 (7.39)

Grade 8

Math						
Female interaction effect	13.49 (21.39)	8.40 (16.30)	1.59 (17.09)	-6.70 (7.14)	-5.42 (5.58)	-6.75 (5.59)
Male interaction effect	10.37 (21.69)	6.79 (17.96)	9.99 (19.32)	11.85 (7.31)	9.44 (6.15)	11.46* (6.18)
Science						
Female interaction effect	1.91 (13.07)	5.16 (10.10)	3.37 (9.46)	-9.42 (6.90)	-6.73 (4.99)	-5.68 (5.21)
Male interaction effect	25.41* (13.36)	16.70 (11.04)	18.53* (11.01)	11.17 (7.15)	10.08 (5.42)	8.04 (5.55)

Regressions of student test scores on teacher's sex separate by student's sex. Coefficients of interaction effect between same sex of students and teachers are presented. Cluster robust standard errors in parentheses.

Specification (1) includes no covariates, (2) includes student background variables age in months, one parent born abroad, speaking another language at home, four categories of numbers of books at home and household size, (3) includes student background variables and teacher's experience, experience squared, teacher having a major in the subject (only available for TIMSS 2003) and two categories of community size. Sample size is identical for all three specifications.

Significance levels: *** 1 percent, ** 5 percent, * ten percent.

Table 4: Gender interaction effects with TIMSS grade 8 gain score

		England			USA		
		(1)	(2)	(3)	(1)	(2)	(3)
TIMSS 1995							
Female	Maths	-5.94 (6.63)	-4.15 (7.07)	-2.65 (7.18)	4.96 (3.46)	4.42 (3.26)	5.38* (3.19)
	Science	4.02 (7.37)	4.33 (6.91)	3.88 (6.85)	-6.81** (3.36)	-3.89 (3.45)	-4.09 (3.55)
Male	Maths	-8.59 (6.22)	1.00 (6.56)	1.19 (6.38)	-9.50** (4.28)	-7.55* (4.10)	-7.37* (4.04)
	Science	5.91 (6.87)	4.44 (5.82)	2.57 (5.63)	4.13 (3.49)	.34 (3.40)	1.78 (3.51)
Joint	Maths	-7.46 (4.57)	-1.64 (4.63)	-.76 (4.56)	-2.59 (1.78)	-1.78 (1.58)	-1.48 (1.47)
	Science	5.13 (5.19)	4.40 (4.58)	4.65 (4.74)	-1.22 (1.68)	-1.74 (1.38)	-1.75 (1.35)
TIMSS 1999							
Female	Maths	7.52 (5.79)	9.19 (6.38)	9.60 (6.32)	1.72 (3.18)	.87 (3.21)	.85 (3.18)
	Science	-4.22 (5.30)	3.93 (5.36)	5.45 (5.18)	-3.02 (3.49)	-2.01 (3.35)	-1.46 (3.30)
Male	Maths	-2.50 (5.86)	1.51 (5.43)	-1.62 (5.41)	4.92 (3.39)	4.07 (3.62)	3.87 (3.53)
	Science	12.08** (5.67)	10.71* (6.10)	6.21 (5.68)	-3.20 (3.45)	-2.17 (3.53)	-2.65 (3.26)
Joint	Maths	3.02 (3.50)	5.78 (3.65)	4.87 (3.72)	3.33** (1.52)	2.47* (1.35)	2.25* (1.33)
	Science	3.36 (3.14)	6.67** (3.35)	6.79** (3.07)	-3.13** (1.41)	-2.11 (1.31)	-1.68 (1.27)
TIMSS 2003							
Female	Maths	10.69** (5.07)	11.59* (6.08)	11.06* (5.54)	-.52 (2.37)	-1.94 (2.59)	-2.40 (2.60)
	Science	4.89 (4.04)	4.30 (4.56)	4.29 (4.44)	-.63 (2.31)	-2.33 (2.46)	-2.25 (2.52)
Male	Maths	4.02 (3.77)	6.64 (4.75)	6.89 (4.83)	-1.66 (2.35)	.60 (2.63)	1.08 (2.65)
	Science	-4.00 (4.29)	1.03 (4.65)	-.14 (4.50)	-1.47 (2.30)	.04 (2.50)	.33 (2.52)
Joint	Maths	6.96** (2.85)	8.91*** (2.96)	9.26*** (2.81)	-1.06 (1.17)	-.66 (1.13)	-.63 (1.13)
	Science	-.005 (2.55)	2.39 (2.64)	2.06 (2.53)	-1.02 (1.17)	-1.15 (1.11)	-1.18 (1.11)

Regressions of student test scores on teacher's sex in math and science separate by student's sex and jointly for girls and boys. Coefficients of interaction effects between same sex of students and teachers in each subject are presented. Cluster robust standard errors in parentheses. In specification (1) dependent variable is difference between math and science test scores and no covariates are included. In specification (2) dependent variable is math test score and science test score is included as covariate. Specification (3) includes for both math and science the covariates teacher's experience, experience squared and teacher having a major in the subject (only available for 2003). For the joint estimations, all specifications include a dummy for gender and specification (3) includes teacher's gender in both subjects and an interaction term between gender and science score. Sample size is identical for all specifications.

Significance levels: *** 1 percent, ** 5 percent, * ten percent.

Table 5: Gender interaction effects with TIMSS grade 8 gain score, exclude fully separated classes

		TIMSS 1995		TIMSS 1999		TIMSS 2003	
		ENG	USA	ENG	USA	ENG	USA
Female	Maths	-2.77 (7.24)	5.43 (3.19)	10.26** (4.86)	1.00 (3.23)	3.40 (5.23)	-2.43 (2.60)
	Science	1.62 (7.66)	-4.22 (3.58)	2.22 (5.24)	-1.39 (3.33)	2.58 (4.58)	-2.28 (2.53)
Male	Maths	-1.39 (6.39)	-7.61* (4.22)	-6.99 (5.37)	2.43 (3.49)	4.36 (4.59)	1.37 (2.64)
	Science	.54 (5.99)	1.74 (3.55)	2.33 (5.90)	-2.79 (3.33)	-4.58 (4.99)	.51 (2.54)
Joint	Maths	-2.84 (4.50)	-1.60 (1.52)	2.71 (2.85)	1.63 (1.27)	4.67** (1.85)	-.49 (1.12)
	Science	2.17 (5.16)	-1.88 (1.36)	3.21 (2.88)	-1.68 (1.24)	-.84 (2.30)	-1.07 (1.09)

All math and science classes that are fully separated by gender are excluded from the sample. Regressions of student test scores on teacher's sex in math and science separate by student's sex and jointly for girls and boys. Coefficients of interaction effects between same sex of students and teachers in each subject are presented. Cluster robust standard errors in parentheses. Dependent variable is math test score. Covariates are science test score, and for both math and science teacher's experience, experience squared and teacher having a major in the subject (only available for 2003). For the joint estimations, all specifications include a dummy for gender and teacher's gender in both subjects and an interaction term between gender and science score. Same specification as specification (3) in Table 4.

Significance levels: *** 1 percent, ** 5 percent, * ten percent.

Annex A - The UK School Public Examination System.

The General Certificate of Secondary Education (GCSE) is taken by all secondary school pupils in England usually at age 15-16 in Year 11. Virtually all candidates take GCSEs in English, Mathematics and Science and up to around 7 other subjects after a two year course. All candidates gain grades from A*, A down to G. Grades A* to E are pass grades. In addition to examinations sat in May/June most subjects have coursework components which usually contribute 20-40% of the final GCSE grade.

Five different exam boards offer GCSE exams and all these boards are under the control of the Joint Council for Qualifications who deal with examination conduct and legislation. These boards are self-sufficient non-profit making organisations.

GCSEs are now seen as Key Stage 4 exams, where the preceding Key Stage 1, 2 and 3 tests are sat in core subjects by pupils aged 7, 11 and 13 respectively. These Key Stages are part of a National Curriculum which is prescribed by the Department for Education and Skills.

The national GCSE system was introduced in September 1986 (with the first exams being sat in May 1987). The GCSE system replaced a two-tier system of O-levels (Ordinary levels) and CSE (Certificate in Secondary Education) exams which were typically sat by pupils in Grammar schools and Secondary Modern Schools. The O-level exams did not have a course work component.

Table A1: Summary statistics

	MATH/READING				SCIENCE			
	England		USA		England		USA	
	all	sample	all	sample	all	sample	all	sample
<i>Grade 4</i>								
PIRLS (Reading)								
observations	3,156	2,534	3,763	2,998				
Female student	0.52	0.52	0.51	0.51				
Female teacher	0.80	0.80	0.88	0.88				
Female student *	0.42	0.43	0.45	0.45				
female teacher								
Male student *	0.11	0.11	0.06	0.06				
male teacher								
<i>Grade 8</i>								
TIMSS 1995								
observations	3,579	1,199	10,973	7,757	3,579	1,119	10,973	6,969
Female student	0.47	0.48	0.50	0.50	0.47	0.47	0.50	0.51
Female teacher	0.46	0.45	0.69	0.71	0.39	0.39	0.57	0.59
Female student *	0.23	0.23	0.35	0.36	0.21	0.21	0.29	0.31
female teacher								
Male student *	0.31	0.31	0.16	0.15	0.36	0.34	0.21	0.21
male teacher								
<i>Grade 8</i>								
TIMSS 1999								
observations	2,960	1,510	9,072	6,196	2,960	1,693	9,072	6,065
Female student	0.49	0.49	0.50	0.51	0.49	0.50	0.50	0.51
Female teacher	0.48	0.46	0.60	0.63	0.42	0.44	0.48	0.47
Female student *	0.27	0.26	0.30	0.32	0.23	0.24	0.24	0.24
female teacher								
Male student *	0.30	0.31	0.19	0.18	0.30	0.30	0.27	0.27
male teacher								
<i>Grade 4</i>								
TIMSS 2003								
Observations	3,585	1,879	9,829	6,903	3,585	1,888	9,829	6,890
Female student	0.50	0.51	0.50	0.50	0.50	0.51	0.50	0.50
Female teacher	0.74	0.75	0.85	0.87	0.74	0.75	0.85	0.87
Female student *	0.36	0.38	0.42	0.44	0.36	0.38	0.42	0.44
female teacher								
Male student *	0.14	0.13	0.08	0.06	0.14	0.13	0.08	0.06
male teacher								
<i>Grade 8</i>								
TIMSS 2003								
Observations	2,830	1,350	8,912	6,709	2,830	1,097	8,912	6,564
Female student	0.50	0.48	0.52	0.52	0.50	0.49	0.52	0.52
Female teacher	0.54	0.57	0.65	0.64	0.55	0.48	0.54	0.54
Female student *	0.30	0.30	0.34	0.34	0.29	0.26	0.27	0.28
female teacher								
Male student *	0.27	0.25	0.17	0.17	0.25	0.29	0.22	0.22
male teacher								

Average of the student population, weighted by students' sampling probability. PIRLS 2001, TIMSS 2003. For TIMSS at grade 4, there is no distinction between math and science teachers. "All" comprises all sampled students available in the data. Number of observations for teacher's gender may be lower due to missing values.

“Sample” comprises only students for which information on student background and most important class and school characteristics is available, same as in Table 3. “Male student * male teacher” states which fraction of all students is male and has a male teacher.

Table A2: Average scores by subject, country and gender in sample, TIMSS

	MATH			SCIENCE			GAIN SCORE		
	Average score (Std. dev.)		Differen ce (Std. err.)	Average score (Std. dev.)		Differen ce (Std. err.)	Average score (Std. dev.)		Differen ce (Std. err.)
	Female	Male		Female	Male		Female	Male	
TIMSS 1995									
<i>Grade 8</i>									
England	419.08 (80.52)	506.56 (87.48)	15.48*** (4.87)	524.08 (83.93)	545.87 (95.29)	21.79*** (5.39)	-30.30 (51.59)	-43.38 (59.21)	-13.08*** (4.09)
USA	481.29 (79.23)	488.65 (83.53)	7.36*** (1.85)	497.06 (88.96)	516.08 (93.35)	19.02*** (2.18)	-12.47 (52.84)	-23.31 (55.25)	-10.84*** (1.25)
TIMSS 1999									
<i>Grade 8</i>									
England	503.84 (82.37)	528.09 (82.37)	24.25*** (4.16)	538.27 (80.74)	573.65 (85.57)	35.38*** (4.04)	-34.60 (45.01)	-47.13 (46.05)	-12.53*** (2.29)
USA	498.00 (77.68)	508.72 (83.60)	10.72*** (2.05)	505.21 (83.19)	530.18 (90.90)	24.96*** (2.24)	-5.47 (43.80)	-18.22 (45.84)	-12.75*** (1.06)
TIMSS 2003									
<i>Grade 4</i>									
England	532.08 (81.06)	533.39 (84.79)	-1.31 (3.83)	544.44 (74.66)	541.38 (80.38)	-3.06 (3.57)			
USA	511.84 (70.50)	518.93 (74.87)	7.09*** (1.75)	531.15 (73.65)	535.59 (78.64)	4.44** (1.84)			
<i>Grade 8</i>									
England	522.05 (80.20)	524.46 (81.32)	2.41 (4.40)	559.94 (72.08)	576.01 (77.34)	16.07*** (4.53)	-37.76 (44.37)	-52.47 (45.89)	-14.72*** (2.89)
USA	505.50 (72.78)	512.92 (78.24)	6.42*** (1.83)	523.60 (73.15)	540.96 (77.12)	17.36*** (1.86)	-17.49 (45.85)	-28.50 (47.46)	-11.01*** (1.14)

Average math and science scores for female and male students and difference between male and female students in sample used for estimations in Table 3. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

Table A3: Average scores by country and gender in sample, PIRLS

	Average score (Std. dev.)		Difference (Std. err.)
	Female	Male	
England	569.08 (79.28)	548.16 (82.12)	-20.92^{***} (3.21)
USA	560.17 (72.76)	543.82 (80.74)	-16.34^{***} (2.81)

Average reading scores for female and male students and difference between male and female students in working sample. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

Table A4: Coefficients (std. err.) for reading in grade four, PIRLS

	England		USA	
	Male	Female	Male	Female
Age	34.70 ^{***} (7.81)	26.72 ^{***} (6.42)	-13.42 ^{**} (6.27)	-9.35 ^{**} (4.22)
Parent born abroad	9.67 [*] (5.03)	-3.60 (4.34)	-6.24 (5.48)	-10.12 ^{**} (4.27)
Other language	-32.67 ^{***} (7.86)	-39.84 ^{***} (8.17)	-35.94 ^{***} (6.77)	-30.82 ^{***} (7.16)
11-20	26.43 ^{***} (9.49)	40.22 ^{**} (11.29)	41.25 ^{***} (9.77)	36.60 ^{***} (9.59)
21-100	58.71 ^{**} (8.99)	76.90 ^{**} (11.23)	65.67 ^{***} (8.44)	60.48 ^{**} (8.97)
101-200	86.05 ^{***} (9.83)	99.69 ^{**} (12.70)	77.50 ^{***} (9.01)	80.66 ^{***} (9.20)
> 200 books	85.07 ^{***} (9.58)	96.12 ^{***} (12.27)	73.32 ^{***} (9.90)	71.44 ^{***} (10.28)
Household- size	-8.15 ^{***} (1.67)	-7.59 ^{***} (1.72)	-6.03 ^{***} (1.33)	-5.58 ^{***} (1.22)
Teacher's experience	2.51 (1.80)	4.06 ^{**} (1.85)	0.79 (1.39)	1.63 (1.21)
Teacher's experience sq.	-0.09 (0.08)	-0.17 [*] (0.09)	-0.003 (0.05)	-0.05 (0.04)
City	5.22 (7.82)	15.04 (10.03)	-13.40 [*] (7.58)	-11.20 [*] (6.52)
Village	8.89 (7.11)	13.66 [*] (8.16)	12.04 (13.68)	12.35 (14.82)
Gender interaction	-6.71 (7.79)	4.85 (10.57)	-19.15 (12.60)	-0.80 (9.08)
R ²	0.1942	0.1829	0.2031	0.2026
F-stat	16.63	18.23	21.15	18.85
Observations	1254	1280	1473	1525

Regressions of student test scores on interaction effect between same sex of students and teachers and covariates separate by student's sex. Cluster robust standard errors in parentheses. Same regressions as in column (3) in Table 3. Significance level: *** 1 percent, ** 5 percent, * ten percent.

Table A5: Coefficients (std. err.) for math and science in grade four, TIMSS 2003

	England				USA			
	Math		Science		Math		Science	
	Male	Female	Male	Female	Male	Female	Male	Female
Age	21.24 ^{**} (10.02)	39.74 ^{***} (9.23)	19.83 ^{**} (8.85)	38.57 ^{***} (7.69)	-15.55 ^{***} (3.13)	-10.59 ^{***} (3.25)	-9.11 ^{***} (3.31)	-7.80 ^{**} (3.24)
Parent born abroad	-4.11 (8.42)	-25.02 ^{***} (6.59)	-12.42 [*] (7.29)	-22.98 ^{***} (6.62)	-15.01 ^{***} (3.98)	-10.80 ^{***} (3.65)	-19.63 ^{***} (4.07)	-14.79 ^{***} (3.19)
Other language	-39.37 ^{***} (12.73)	-26.86 ^{**} (12.33)	-42.28 ^{***} (12.84)	-24.47 ^{**} (10.08)	-28.34 ^{***} (5.36)	-32.34 ^{***} (4.55)	-32.81 ^{***} (5.38)	-42.56 ^{***} (4.40)
11-20	31.94 ^{***} (11.15)	13.97 (13.58)	29.10 ^{***} (9.28)	18.15 (14.71)	18.15 ^{**} (3.81)	18.77 ^{***} (4.95)	12.56 ^{**} (4.57)	12.50 ^{**} (4.96)
21-100	59.32 ^{***} (10.07)	46.41 ^{***} (11.63)	53.09 ^{***} (9.11)	41.00 ^{***} (12.91)	46.79 ^{**} (3.65)	39.46 ^{***} (5.36)	45.27 ^{***} (4.21)	34.04 ^{***} (4.97)
101-200	78.76 ^{***} (10.88)	56.03 ^{***} (12.84)	74.84 ^{***} (9.49)	52.97 ^{***} (14.13)	64.33 ^{***} (4.98)	57.82 ^{***} (5.96)	66.13 ^{***} (4.91)	51.06 ^{***} (5.41)
> 200 books	92.66 ^{***} (11.38)	84.30 ^{***} (15.03)	90.54 ^{***} (9.29)	80.12 ^{***} (15.57)	59.30 ^{**} (5.40)	59.28 ^{***} (6.22)	54.72 ^{***} (5.75)	56.88 ^{***} (5.86)
Household-size	-0.53 (2.19)	-5.68 ^{***} (1.70)	-2.75 [*] (1.60)	-7.13 ^{***} (1.74)	-4.63 ^{***} (0.98)	-4.49 ^{***} (1.01)	-5.10 ^{***} (1.11)	-5.36 ^{***} (0.94)
Teacher's experience	1.56 (1.90)	1.75 (1.76)	1.20 (1.71)	1.71 (1.49)	1.31 [*] (0.72)	0.88 (0.72)	1.05 (0.77)	0.66 (0.75)
Teacher's experience sq.	-0.04 (0.06)	-0.02 (0.05)	-0.02 (0.05)	-0.03 (0.04)	-0.03 [*] (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.01 (0.02)
Major in subject	22.91 (17.03)	-2.74 (9.98)	13.75 (14.76)	20.56 [*] (11.77)	-8.12 (7.33)	-5.96 (6.85)	-5.65 (6.77)	-4.36 (4.82)
City	-23.57 (11.71)	-18.33 [*] (9.40)	-25.93 ^{***} (9.58)	-13.99 [*] (7.96)	-9.67 (5.92)	-16.96 ^{***} (5.57)	-21.83 ^{***} (6.30)	-22.10 ^{***} (5.53)
Village	11.99 (16.39)	8.82 (11.40)	15.65 (11.38)	5.59 (7.85)	-3.41 (5.31)	-5.65 (4.99)	-1.12 (5.56)	1.65 (5.06)
Gender interaction	8.65 (10.50)	-0.50 (9.19)	1.99 (8.95)	-8.81 (8.11)	2.57 (7.08)	-8.60 (6.19)	-0.23 (7.39)	-7.92 (6.37)
R ²	0.2029	0.2364	0.2447	0.2492	0.1941	0.1943	0.2203	0.2298
F-stat	10.22	18.93	12.51	16.64	28.80	31.04	34.67	41.64
Observations	923	956	927	961	3440	3463	3431	3459

Regressions of student test scores on interaction effect between same sex of students and teachers and covariates separate by student's sex. Cluster robust standard errors in parentheses. Same regressions as in column (3) in Table 3. Significance level: *** 1 percent, ** 5 percent, * ten percent.

Table A6: Coefficients (std. err.) for math and science in grade eight, TIMSS 1995

	England				USA			
	Math		Science		Math		Science	
	Male	Female	Male	Female	Male	Female	Male	Female
Grade 7	-	-	-	-	-48.18 ^{***}	-42.86 ^{***}	-42.80 ^{***}	-44.94 ^{***}
					(5.78)	(4.88)	(5.56)	(5.73)
Grade 9	4.98	36.34 ^{***}	3.14	11.72	-	-	-	-
	(10.64)	(12.53)	(13.14)	(12.33)				
Age	17.48 [*]	2.46	20.82 ^{**}	15.48	-24.02 ^{***}	-21.43 ^{***}	-19.88 ^{***}	-18.08 ^{***}
	(9.02)	(10.64)	(9.51)	(11.25)	(3.51)	(3.11)	(3.73)	(4.03)
Parent born abroad	19.46 [*]	29.23 ^{***}	5.67	29.79 ^{***}	2.56	4.07	-1.55	-2.03
	(10.02)	(10.65)	(6.50)	(11.26)	(5.62)	(4.69)	(5.68)	(4.86)
Other language	-33.40	-8.97	-33.16 [*]	-53.18 ^{**}	-22.12 ^{**}	-18.71 ^{***}	-41.08 ^{**}	-38.49 ^{**}
	(22.12)	(34.09)	(17.40)	(22.48)	(5.48)	(6.60)	(7.26)	(6.03)
11-20	46.27	25.45 [*]	39.41 ^{**}	46.36 [*]	13.81 ^{***}	19.92 ^{***}	10.97	24.91 ^{***}
	(18.46)	(14.10)	(16.35)	(24.60)	(5.78)	(6.01)	(7.55)	(6.91)
21-100	85.34 ^{***}	59.37 ^{**}	68.41 ^{***}	86.67 ^{**}	49.47 ^{**}	49.59 ^{***}	54.43 ^{**}	69.05 ^{**}
	(18.12)	(16.90)	(16.52)	(25.07)	(5.39)	(6.16)	(6.48)	(6.47)
101-200	101.60 ^{***}	84.42 ^{***}	106.41 ^{***}	120.19 ^{***}	73.02 ^{***}	73.43 ^{***}	77.29 ^{**}	91.37 ^{***}
	(18.01)	(14.62)	(16.47)	(26.91)	(6.52)	(6.74)	(7.17)	(7.16)
> 200 books	128.86 ^{***}	107.64 ^{***}	143.35 ^{***}	147.91 ^{***}	79.55 ^{**}	88.07 ^{***}	89.32 ^{***}	116.95 ^{***}
	(18.75)	(15.76)	(17.71)	(26.01)	(6.65)	(7.35)	(7.96)	(7.59)
Household-size	-8.11	-10.69 ^{***}	-9.01 ^{***}	-9.22 ^{***}	-4.08 ^{***}	-4.34 ^{***}	-5.92 ^{***}	-6.40 ^{***}
	(2.97)	(2.88)	(3.32)	(2.58)	(1.00)	(0.86)	(1.09)	(1.04)
Teacher's experience	-0.39	0.07	0.66	-0.62	0.96	-0.07	1.56	0.63
	(1.58)	(1.70)	(1.54)	(1.17)	(1.24)	(1.10)	(1.02)	(0.90)
Teacher's experience sq.	-0.02	0.01	0.01	0.05	0.01	0.03	-0.03	-0.01
	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)
City	-36.19 ^{**}	-19.17	-18.94	-14.12	-8.83	-8.26	-4.90	-7.26
	(17.58)	(11.97)	(17.40)	(10.79)	(8.74)	(8.07)	(10.34)	(9.03)
Village	-8.75	-0.91	-8.15	8.21	-16.13 ^{**}	-7.65	-0.59	5.11
	(10.74)	(13.03)	(12.12)	(19.83)	(8.06)	(7.54)	(9.83)	(9.10)
Gender interaction	11.89	15.97 [*]	4.15	10.75	-6.06	2.42	-10.70	6.33
	(11.34)	(8.37)	(11.62)	(7.89)	(6.20)	(6.13)	(6.77)	(6.28)
R ²	0.2549	0.2687	0.3036	0.3091	0.2211	0.2178	0.2174	0.2718
Observations	624	575	590	529	3793	3964	3386	3583

Regressions of student test scores on interaction effect between same sex of students and teachers and covariates separate by student's sex. Cluster robust standard errors in parentheses. Same regressions as in column (3) in Table 3. Significance level: ^{***} 1 percent, ^{**} 5 percent, ^{*} ten percent.

Table A7: Coefficients (std. err.) for math and science in grade eight, TIMSS 1999

	England				USA			
	Math		Science		Math		Science	
	Male	Female	Male	Female	Male	Female	Male	Female
Age	3.08 (11.57)	9.63 (10.14)	7.11 (8.78)	22.04** (9.05)	-21.88*** (3.48)	-16.66*** (3.48)	-16.42*** (3.16)	-11.05*** (3.38)
Parent born abroad	18.16** (7.61)	19.23** (9.14)	11.37 (8.50)	14.01* (8.22)	-16.48*** (5.40)	-7.14 (5.35)	-20.32*** (6.16)	-13.96*** (5.22)
Other language	2.94 (23.49)	-22.70 (17.06)	-18.36 (16.65)	-36.19*** (11.47)	-22.61*** (6.28)	-18.19*** (6.38)	-35.11*** (6.92)	-21.32*** (6.85)
11-20	23.18 (15.14)	25.12 (15.34)	19.74 (14.49)	-2.77 (13.79)	28.27*** (6.21)	20.32*** (6.98)	35.40*** (6.70)	30.22*** (8.12)
21-100	44.35*** (11.53)	56.40*** (13.30)	56.43*** (12.26)	44.36*** (12.31)	47.57*** (5.84)	46.96*** (6.74)	62.19*** (6.66)	55.35*** (8.03)
101-200	60.59*** (13.91)	76.24*** (14.53)	74.40*** (14.24)	71.18*** (13.46)	71.22*** (6.07)	69.90*** (9.54)	83.29*** (7.24)	80.04*** (7.84)
> 200 books	91.49*** (14.69)	110.01*** (16.00)	110.30*** (14.15)	109.01*** (13.73)	81.57*** (6.59)	85.69*** (7.53)	98.15*** (6.56)	104.16*** (8.12)
Household- size	-8.31** (2.15)	-4.31** (1.91)	-11.22*** (2.27)	-5.21** (2.01)	-4.24*** (0.74)	-4.63*** (0.66)	-6.79*** (0.83)	-6.48*** (0.79)
Teacher's experience	-2.54 (1.86)	-1.71 (1.65)	1.98 (1.79)	1.24 (1.85)	1.12 (1.21)	2.34** (1.00)	-1.42 (1.24)	-0.92 (1.02)
Teacher's experience sq.	0.09* (0.05)	0.06 (0.05)	-0.06 (0.06)	-0.01 (0.06)	-0.01 (0.03)	-0.06** (0.03)	0.05 (0.04)	0.03 (0.03)
City	36.68*** (12.11)	0.26 (14.48)	15.35 (11.12)	0.01 (11.35)	0.27 (9.65)	0.48 (7.81)	-9.24 (8.76)	-14.01* (7.83)
Village	61.27*** (16.28)	24.03 (16.76)	53.78*** (14.30)	29.81** (11.93)	-17.33* (9.04)	-7.29 (8.39)	-19.08** (8.78)	-15.39* (9.25)
Gender interaction	10.06 (10.24)	12.80 (12.10)	1.00 (8.42)	17.03** (8.15)	3.92 (8.37)	-4.13 (6.16)	6.35 (6.83)	-1.76 (6.33)
R ²	0.2682	0.1843	0.3037	0.2979	0.1921	0.1868	0.2179	0.2131
F-stat	12.15	10.73	15.22	15.73	19.73	24.75	30.47	38.02
Observations	737	773	814	879	2981	3215	2908	3157

Regressions of student test scores on interaction effect between same sex of students and teachers and covariates separate by student's sex. Cluster robust standard errors in parentheses. Same regressions as in column (3) in Table 3. Significance level: *** 1 percent, ** 5 percent, * ten percent.

Table A8: Coefficients (std. err.) for math and science in grade eight, TIMSS 2003

	England				USA			
	Math		Science		Math		Science	
	Male	Female	Male	Female	Male	Female	Male	Female
Age	5.80 (9.43)	19.08 (11.04)	3.38 (6.17)	5.49 (6.95)	-17.77 ^{***} (3.39)	-14.29 ^{**} (3.27)	-10.62 ^{***} (3.24)	-9.83 ^{***} (3.46)
Parent born abroad	-0.72 (14.10)	4.90 (13.12)	4.69 (11.59)	6.08 (8.53)	-3.12 (5.08)	-5.96 (4.36)	-10.97 ^{**} (4.42)	-10.46 ^{**} (4.43)
Other language	-10.22 (18.46)	-44.45 ^{**} (18.33)	-11.97 (17.07)	-7.43 (33.63)	-9.76 (7.05)	-19.12 ^{***} (6.23)	-18.79 ^{***} (7.26)	-26.96 ^{***} (5.66)
11-20	45.70 ^{***} (11.68)	42.98 ^{***} (10.96)	45.10 ^{***} (12.60)	39.58 ^{**} (17.56)	25.83 ^{***} (4.04)	18.35 ^{***} (4.55)	21.19 ^{***} (4.51)	20.42 ^{***} (4.72)
21-100	62.97 ^{***} (12.69)	69.01 ^{***} (12.08)	62.66 ^{***} (12.28)	51.94 ^{***} (17.72)	57.55 ^{***} (4.38)	45.85 ^{***} (4.63)	57.29 ^{***} (4.33)	47.74 ^{***} (4.71)
101-200	74.37 ^{***} (12.82)	93.00 ^{***} (14.22)	78.07 ^{***} (17.48)	90.32 ^{***} (19.78)	76.40 ^{***} (5.36)	67.88 ^{***} (5.27)	79.52 ^{***} (5.52)	76.05 ^{***} (5.64)
> 200 books	96.83 ^{***} (14.11)	112.31 ^{***} (13.74)	109.01 ^{***} (14.35)	107.71 ^{***} (19.78)	82.48 ^{***} (6.09)	84.60 ^{***} (5.90)	90.76 ^{***} (5.57)	93.90 ^{***} (5.60)
Household- size	-3.17 (2.10)	-6.83 ^{**} (2.51)	-6.16 ^{**} (2.37)	-4.07 (2.54)	-3.94 ^{***} (1.16)	-3.22 ^{***} (1.07)	-5.85 ^{***} (1.04)	-5.87 ^{***} (0.96)
Teacher's experience	-1.05 (2.41)	1.08 (2.05)	-0.95 (1.96)	-0.66 (1.55)	2.06 ^{**} (1.04)	1.72 [*] (0.94)	1.63 (0.98)	0.92 (0.98)
Teacher's experience sq.	0.03 (0.07)	-0.04 (0.05)	0.05 (0.06)	0.03 (0.04)	-0.05 [*] (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.02 (0.03)
Major in subject	13.21 (13.97)	42.40 ^{**} (16.29)	14.48 (10.78)	8.65 (10.70)	9.59 (5.94)	7.71 (5.34)	3.96 (5.63)	2.20 (5.14)
City	-20.06 (20.94)	-14.17 (18.08)	-6.27 (16.80)	-1.91 (11.00)	-5.56 (9.42)	-5.70 (8.26)	-14.50 [*] (8.32)	-11.78 (7.47)
Village	-12.26 (18.72)	-14.71 (18.19)	-26.92 (19.85)	-28.17 (17.22)	-4.01 (7.15)	-2.53 (6.07)	3.04 (6.68)	2.86 (5.64)
Gender interaction	9.99 (19.32)	1.59 (17.09)	18.53 [*] (11.01)	3.37 (9.46)	11.46 [*] (6.18)	-6.75 (5.59)	8.04 (5.55)	-5.68 (5.21)
R ²	0.1618	0.3199	0.2468	0.2931	0.1875	0.2044	0.2457	0.2661
F-stat	6.01	12.72	8.35	7.63	21.66	28.12	26.28	31.96
Observations	712	638	572	525	3190	3519	3114	3450

Regressions of student test scores on interaction effect between same sex of students and teachers and covariates separate by student's sex. Cluster robust standard errors in parentheses. Same regressions as in column (3) in Table 3. Significance level: *** 1 percent, ** 5 percent, * ten percent.

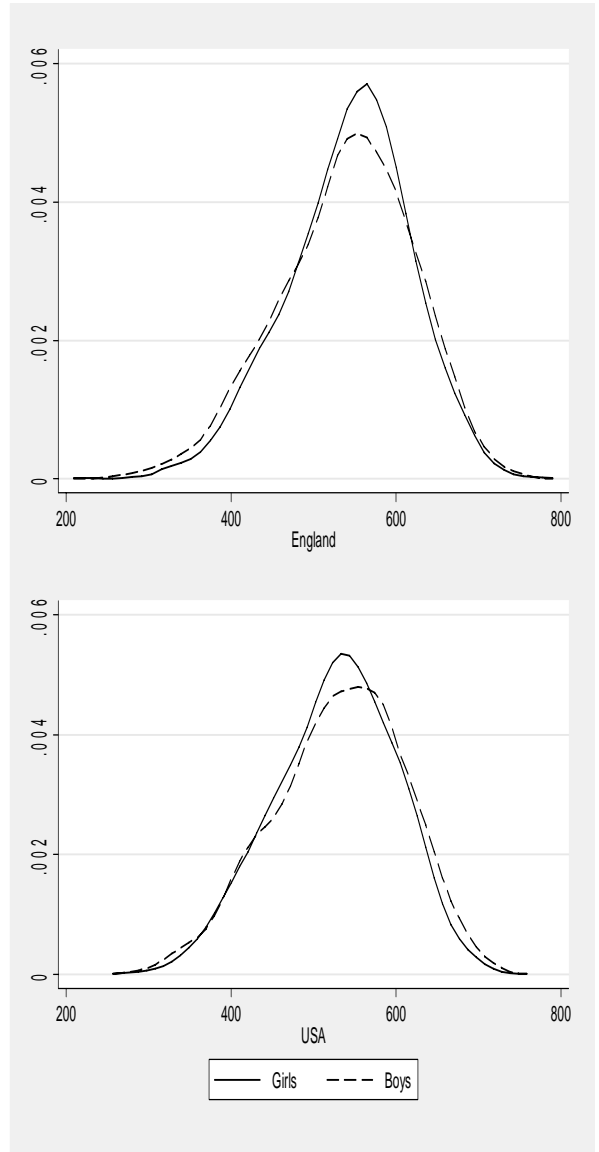
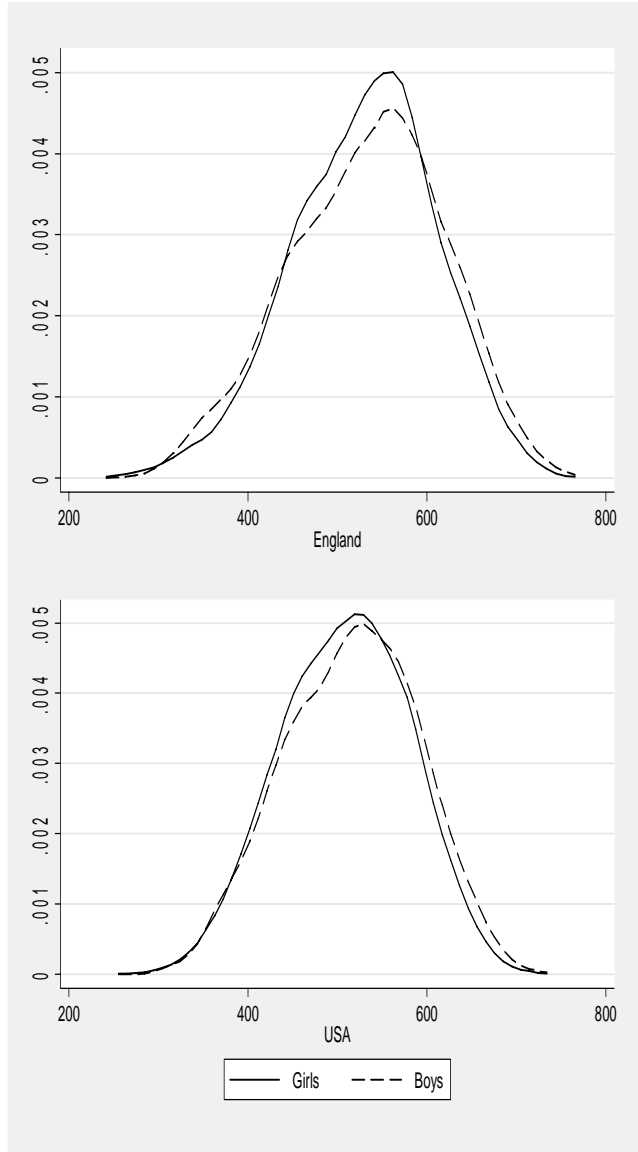
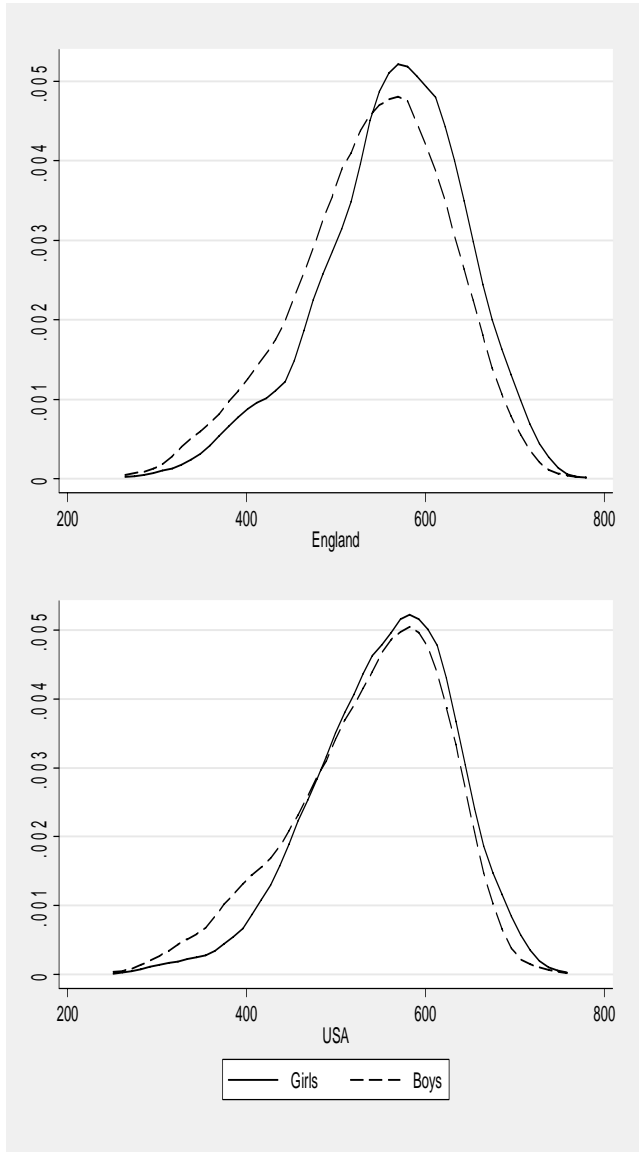
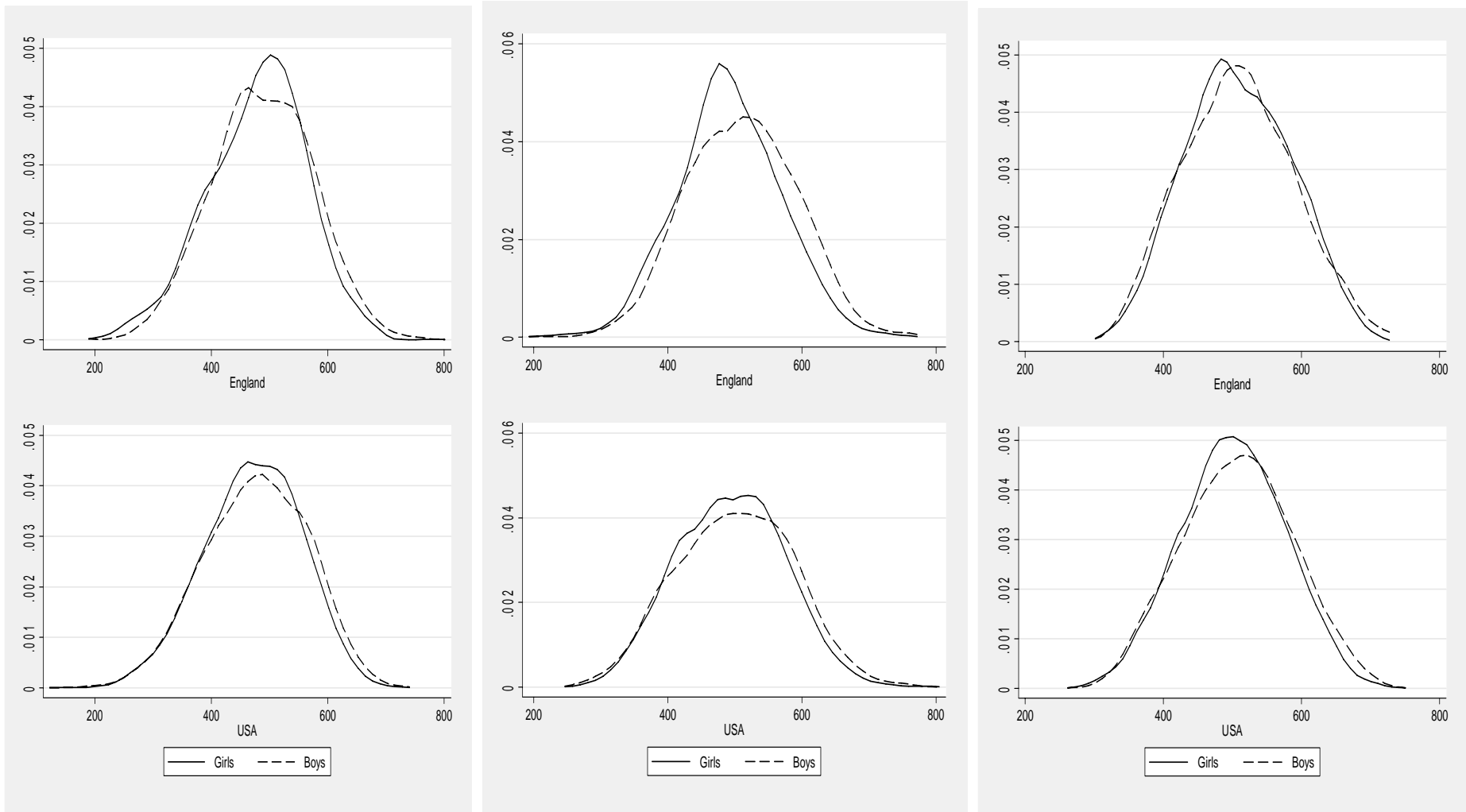


Figure A1: Distribution of scores at grade 4

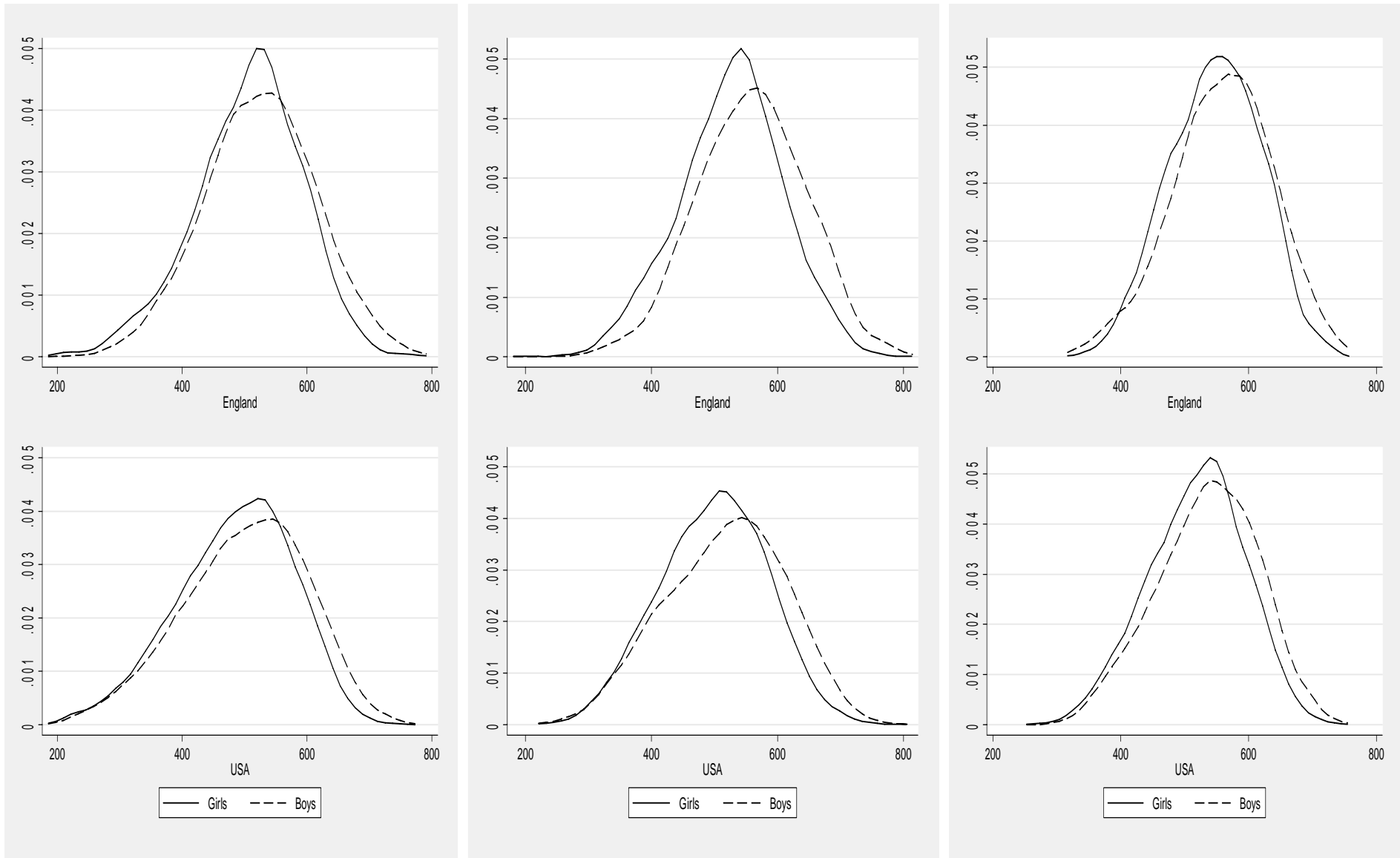


1995

1999

2003

Figure A2: Distribution of TIMSS maths scores at grade 8 by year

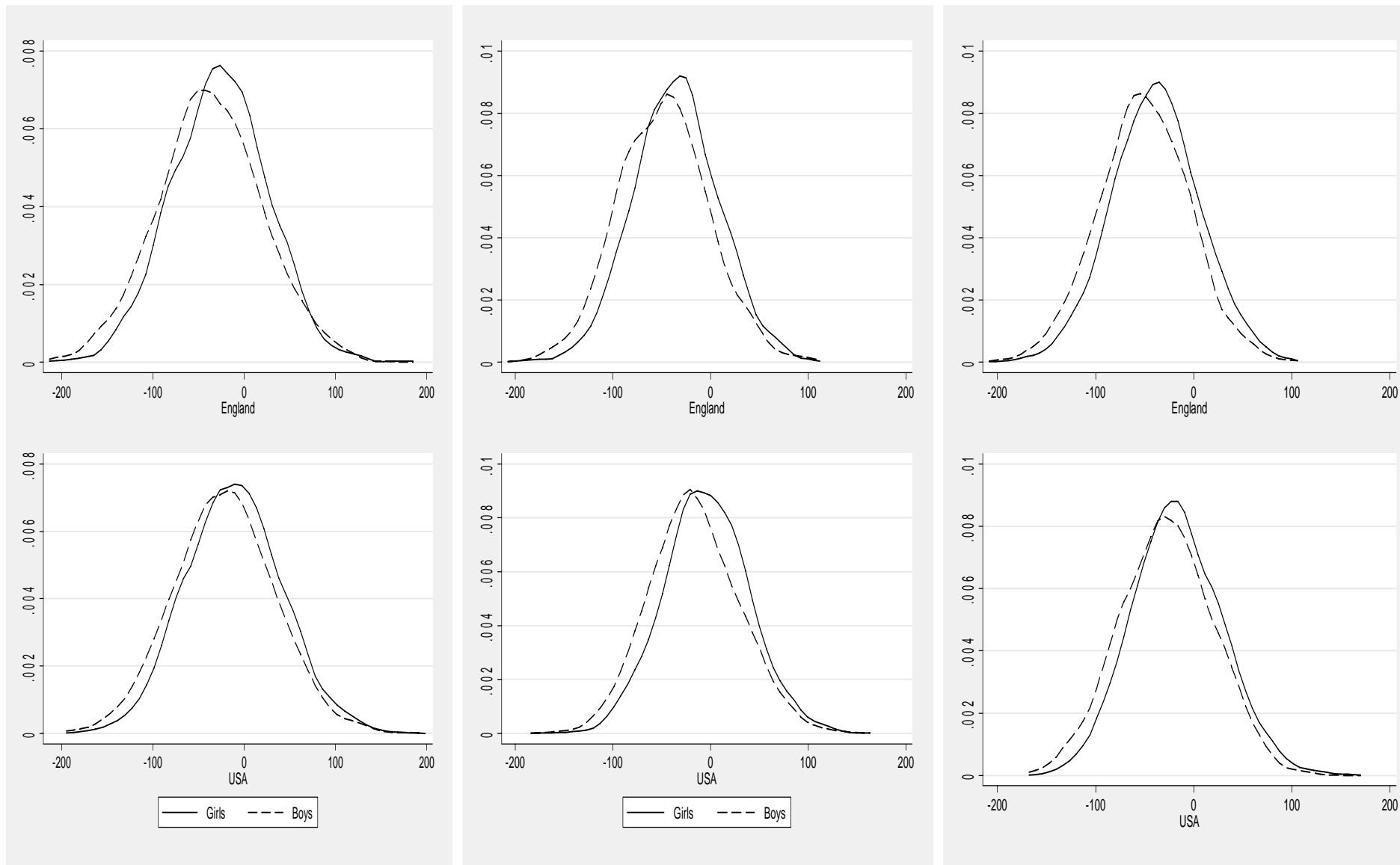


1995

1999

2003

Figure A3: Distribution of TIMSS science scores at grade 8



1995

1999

2003

Figure A4: Distribution of TIMSS GAIN score (Maths-Science) at grade 8