THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

Girl Power? An Analysis of Peer Effects using Changes in the Gender Make-up of the Peer Group

Steven Proud

Revised March 2010
Original version October 2008

Working Paper No. 08/186

Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bristol.ac.uk/cmpo/

Tel: (0117) 33 10799
Fax: (0117) 33 10705
E-mail: cmpo-office@bristol.ac.uk

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

ISSN 1473-625X

CMPO Working Paper Series No. 08/186

Girl Power?
An analysis of peer effects using exogenous changes in the gender make-up of the peer group.

Steven Proud[1]

[1] CMPO, University of Bristol

March 2010
Original Version January 2008

Abstract
The effect of a child's peers has long been regarded as an important factor in affecting their educational outcomes. However, these effects are often difficult to estimate. I use exogenous changes in the proportion of girls within English school cohorts to estimate the effect of a more female peer group, estimated in all schools, and in a subset of schools that only include one classroom per academic year. I find significant negative effects of a more female peer group on boys' outcomes in English. In maths and science, all pupils benefit from a more female peer group up until age 11.

Keywords: peer groups, education

JEL Classification: J1, D1, I21,

Electronic version: www.bristol.ac.uk/cmpo/publications/papers/2008/wp186.pdf

Acknowledgements
The author would like to thank the ESRC for funding support, Professor Simon Burgess, Dr Deborah Wilson, Dr Sarah Smith and Professor Anna Vignoles and participants at CMPO seminars for comments and suggestions.

Address for Correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
Steven.Proud@bristol.ac.uk
www.bris.ac.uk/Depts/CMPO/

CMPO is jointly funded by the Leverhulme Trust and the ESRC

1. Introduction

Perhaps one of the most influential educational reports of the 20th century, The Coleman report (Coleman et al (1966), 86) introduced the idea that "Attributes of other students account for far more variation in the achievement of minority children than do any attributes of school facilities and slightly more than do attributes of staff." The report was commissioned to investigate the level to which school level integration was progressing following the removal of the segregation laws. The conclusions were that schooling increased the disparity between whites and blacks. This led to an interest in the impact of peer groups on educational outcomes. This led to interest in the impact of the make-up of the peer group on educational outcomes. Early studies include Winkler (1975) who found differential effects of the composition of peer groups on different races, and Summers and Wolfe (1977) who found that both black and non-black pupils benefited from a more balanced mix of black and non-black pupils, and also commented that students who tested at or below grade level were helped by being in a school with high achievers.

However, Manski (1993) identifies problems with trying to measure peer group effects. Peer effects can be split into three different types; endogenous effects, correlated effects and exogenous effects. In the case of endogenous effects, decisions made by the individuals within the peer group directly affect the decisions made by other members of the peer group. The second effect is a correlated effect, which is largely due to members of a peer group having some trait in common, which in turn influences the outcomes of the peer group. The final type of effect is an exogenous effect, where one's actions depend on the exogenous characteristics of one's peers. Learning outcomes appear to be endogenous effects, as for instance, a child's desire to work hard could affect other children in the classes' decision on whether to work hard or to misbehave. Manski discusses the problems inherent with such endogenous peer effects when trying to infer the effects that members of a reference group have on its own members (the reflection problem), and argues that it is not possible to make inferences on effects unless one has prior knowledge of the make-up of the reference group. Furthermore, he argues that studies using apparent random distribution may experience bias to the apparent peer effect if there are unseen family characteristics that are in common with the reference group.

This paper builds on work by Hoxby (2000)[1] using exogenous changes in the gender make-up of the within school peer group to estimate the effect of a child's peers on their educational outcomes. Hoxby's initial strategy utilizes the credibly exogenous variation in the distribution of females across cohorts within a grade, using the raw proportion of girls as a measure of the peer group. She then combines this with the test-score gap between girls and boys to estimate the effect of an exogenous change in the ability of the peer group. These positive effects of a more female peer group are backed up by Lavy and Schlosser (2007) who find a positive effect in Israeli schools. Furthermore, Whitmore (2005) finds positive effects of a high proportion of girls in kindergarten through to the second grade on the outcomes of both boys and girls. However, Whitmore (2005) also suggests that in the third grade there is evidence that boys do worse in a class with a high fraction of girls. Hansen et al (2006) find that female dominant and equally mixed groups perform better than male dominated groups.

This paper utilizes the same strategy as Hoxby (2000), but builds upon it in three important ways. First, I take advantage of the fact that in England, there is a legal upper limit of 30 on class sizes for children in infant schools[2]. I use this fact to separate schools that appear to only have one class per cohort to estimate classroom level effects rather than school-level effects. Secondly, I investigate whether there is any bias to the estimates by including a measure of the average socioeconomic status of the male and female pupils separately using the proportion of boys (or girls) who receive free school meals (FSM) within the cohort in the school. Finally, I examine the effect of a more female peer group on the average value added score from one national assessment to the next. This analysis uses data on English pupils from the Pupil Level Annual School Census (PLASC) and the National Pupil Database (NPD). This data includes pupils' results from national assessments and demographics of the pupil, such as age within year, gender, ethnicity and free school meals status. These assessments are Key Stage 1 (KS1) sat at age 7, Key Stage 2 (KS2) sat at age 11, Key Stage 3 (KS3), sat at age 14 and GCSE sat at age 16[3].

I find significant negative effects of a more female peer group for males in English at all levels of assessment, and significant positive effects of a more female peer group on both boys and girls in maths and science, although these effects largely disappear post age 11. These effects

[1] Also published in abridged form as Hoxby (2002)
[2] Infant schools cover ages 4 to 7.
[3] A more detailed description of the English schools system is given in the data section.

all combine to give large and significant average negative effects of ability of the peer group. The omission of the socioeconomic status in the initial models has no significant bias on the coefficient on the proportion of the school-cohort that is female. The value added model shows strong significant positive effects of a more female peer group between ages 7 and 11 in English for both girls and boys, and between ages 11 and 14 for girls in mathematics and science. Furthermore, considering the effect of more females in the class as a proxy for changes in ability, I demonstrate that the magnitudes of the effects are too large, and of the wrong sign, to be explained by small changes in ability.

This paper begins by examining whether there is a credible difference between boys and girls outcomes in order to investigate whether any observed effects are caused by changes in the ability of the peer group. Section 3 discusses the methodology used in this paper. Section 4 examines the PLASC and NPD dataset used here, section 5 discusses the results. Finally, section 6 offers discussion and conclusions.

2. The Gender Gap.

Hoxby (2000) suggests a possible mechanism for her observed effects that differential ability of boys and girls in the classroom could be accountable for these effects. However, the results she finds are too large to be solely explained by changes in the ability of the peer group. Lavy and Schlosser (2007) suggest that their results move in the opposite direction to what would be expected were they caused by differences in the prior attainment of the peer group alone, and also go on to show that the presence of more boys in the classroom is associated with more disruption, causing a negative learning environment.

In order to try and validate these previous studies, I investigate whether the size and direction of effect in this study is credible to be attributable to differences in the ability of the peer group caused by the differential achievement of girls and boys. In order to estimate the effect of a more able peer group using changes in the gender make-up of the class, it is necessary for there to be a significant difference in outcomes for girls and boys in school outcomes.

Hallinan and Sorenson (1987) consider reasons for the differential achievement levels in mathematics, with boys holding the advantage. Whilst they conclude that mathematics teaching within stratified groups does not have a differential effect for girls and boys, they do find that the initial grouping decision is indeed influenced by the gender of the pupil. Male high achievers are far more likely to be assigned to a high achieving group than female high achievers, indicating some unseen factors also affecting the grouping decision (or alternatively just some prejudice against girls in mathematics).

Shibley Hyde et al (1990) carry out a meta-analysis of research on the magnitude of the gender gap in mathematics using 100 separate studies. Whilst they found a male dominance in the subject, this was decreasing over time. Also, some bias was present due to a self-selection problem. When considering performances based on samples of the entire population, females in fact had an advantage albeit a negligible one. However, as samples became more selective, a gender gap became apparent favouring males, which has reduced over time. This implies that either boys are more likely to be the ones to gain the higher grades, or that boys are more likely to drop mathematics if they are not good at it.

Machin and MacNally (2005) examine the education system in England, and show that girls outperform boys in all schools from primary school onwards, particularly in English. In mathematics, the story is a little more complicated, with little difference between the proportion of boys and girls reaching the target grade at age 11. However, they show a clear advantage for girls in English. Similarly, Burgess et al (2004) find that the gender gap is largely seen with girls outperforming boys in English, with very little difference in performance in mathematics and science between equivalent male and female students.

Gorard et al (2001), examining the gender gap in Wales, suggest that all pupils enter education on an equal footing, which is supported by little difference in outcomes between boys and girls at key stage 1. They also suggest that there is little gap between boys and girls in mathematics at key stages 2 and 3, but until recently, there had been a gap in favour of boys at GCSE. For science, they suggest there is little gap between key stages 1 and 3, with a historical gap between girls and boys at GCSE. In English, again, they find little difference at key stage 1, but they find at older levels, there is a much larger proportion of girls achieve grade D or higher.

The majority of the literature suggests there is a significant difference in outcomes for boys and girls in English, suggesting a change in the gender make-up of the classroom would be associated with a change in the average prior attainment of that class. There is also some suggestion of a gender gap in mathematics, but it has moved from an advantage for boys to an insignificant difference in recent years.

3. Methodology

The methodology in this paper uses the same basic methodology used by Hoxby (2000), utilising idiosyncratic changes in the proportion of pupils in the school cohort that are female as a measure of the peer group. This can then be combined with the difference in outcomes associated with the gender of the pupils to try to estimate the effects of a more able peer group on outcomes, and to investigate whether there are more mechanisms in play than simply higher ability peers helping to increase the performance of the rest of the peer group.

I begin with an individual-level educational production function. The model uses the assumption that any school *j* at a given key stage, *g*, has an average outcome for male (female) pupils, which is constant across cohorts, *c*, and differences from this mean can be explained by peer group effects, other factors not correlated with the peer effects and some unobserved random factor. So, for a female pupil *i*, there is a production function thus

$$A_{i,gjc} = \mu_{female,gj} + \gamma_{female} p_{gjc} + \alpha_{female} X_{i,gjc} + \varepsilon_{i,gjc} \qquad (1)$$

where μ is a school-level fixed effect, consisting of a constant, an average school outcome and a school-level fixed effect, *p* is the proportion of pupils in the school-cohort that are female, which is the peer group influence that we are interested in, and *X* represents other pupil-level exogenous and constant variables, which also includes year and key-stage dummies. The dependent variable *A* is the individual's score within the school year. The levels represented are *i* for each individual pupil, *female* (or *male*), *g* representing grade, or exam-level being taken, *j* representing the school attended and *c* representing the cohort the pupil is a member of. This production function assumes that male and female students experience different effects from the proportion of pupils who are female, as well as other exogenous factors.

The exogenous and constant variables, *X*, consist of fixed family background effects (*F*), the pupil's underlying ability (*U*) and various exogenous factors (χ), including year dummies and dummies for the level of the examination.

$$X_{i,gjc} = F_{i,gjc} a + U_{i,gjc} b + \chi_{i,gjc} c + e_{i,gjc} \qquad (2)$$

Since the identification strategy operates at a school level, when taking means, I assume that *F* and *U* are drawn from a population with unchanging demographics. Furthermore, I assume that these effects are uncorrelated with the probability of a child being female, and any time-invariant effects should not bias the effects of a more female peer group.

This individual model (1) can be averaged to a school level average. However, since males and females have different average outcomes, whilst a school average would be directly affected by the proportion of pupils in the school that are female, I use separate specifications for male and female pupils, which will not be affected in this way.

$$A_{female,gjc} = \mu_{female,gj} + \gamma_{female} p_{gjc} + \alpha_{female} X_{female,gjc} + \varepsilon_{female,gjc} \quad (3)$$

The motivation behind this model is that at a given exam, a school has an average outcome that is achieved, and each year there is a variation around this mean, that is influenced by the proportion of pupils that are female and other exogenous effects.

In order to remove the school-level fixed effects, following Hoxby (2000) I take first differences across cohorts within a given key stage,

$$A_{female,gjc} - A_{female,gj(c-1)} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \varepsilon_{female,gjc} - \varepsilon_{female,gj(c-1)} \quad (4)$$

$$\Rightarrow \Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \Delta \varepsilon_{female,gjc} \quad (5)$$

This identification strategy depends on there being no endogenous component of the change in gender make-up of a school. Since the distribution of genders of pupils can be seen as credibly random, then it can be argued that changes in gender makeup should also be credibly random, and as the size of school increases the proportion of girls should tend to the national average.

There is a potential problem with this strategy. Since there is no data on classroom level interactions within the school, it is possible that the magnitude of effect could be mis-estimated. That is there could be non-random allocation of pupils to classrooms within a school, meaning a pupil who attends a large school with a large proportion of girls could still be taught in a male dominated classroom. In order to address this possibility, I use the fact that in England there has been a legal limit placed on the size of infant class sizes (ages 4 to 7) of 30, which was instituted in 2002. This allows me to examine schools with 30 or fewer pupils within the school-year as a proxy for schools that teach their pupils in one class per year. I show later that this can be extended for infant schools for the period before 2002. Whilst there is no such limit imposed on junior schools (serving pupils aged 7 to 11), many junior schools are linked to an infant school, and follow a similar policy with regards classroom allocation. I show in the data section that there is a similar structure of school sizes in junior and infant schools. Thus, I define a small school to be one that has thirty or fewer pupils in every observed cohort, whilst a large school is defined to be one that has more than thirty pupils in every observed cohort. Pupils over the age of 11 are educated in larger schools, and so we cannot extend the strategy further.

In order to try and keep the treatment group constant across the treatment period, I consider only the pupils who stay in the same school between Key Stage 1 and 2 and between key stages 3 and 4. However, the vast majority of children in England change schools between year 6 (key stage 2) and year 7 (key stage 3), and so without any further information about the school attended, I can make the assumption that the pupils are at a fixed school in years 7 to 9, which will be the case for the vast majority of pupils. Thus, for the key stage 2 to 3 measures, I consider those pupils who have moved schools between the exams. The number of pupils who appear in the sample, and the number omitted are shown in table 2.

Since I am not interested in the time or grade effects in the structural model, I simply include year and grade dummies in the first difference equation.

3.1 Tests of robustness.

It is possible that particular schools have policies on admission that makes the proportion of pupils that are female as an endogenous measure, or that variation in the gender makeup of the school follows a non-random pattern due to some other external factor. In order to examine this possibility, I use a similar strategy to Hoxby (2000). That is, for every school within grade, I perform a regression of the proportion of pupils that are female against a linear time trend and a constant. The order of the years within the schools is then randomised, and a further regression is performed, again on a linear (false) time trend the R-squared values from the two regressions are compared. Schools with a ratio of greater than 1.20 for the real time trend R-squared to the false time-trend R-squared are dropped from the sample. Whilst Hoxby (2000) also included non-linear trends, since I only have 3 time observations for GCSE, this is not possible at this level in my data, due to a lack of degrees of freedom. This results in approximately half of the schools being dropped, and a comparison of the results for the sub-sample and the full sample is reported in table 5

Finally, in order to ensure that the linear model of the peer effects is the correct specification, I use a regression including the interaction between the change in the proportion of pupils that are female and the quartile that this is in.

$$\Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \delta_1 \Delta p_{female,gjc} q_1 + \delta_2 \Delta p_{female,gjc} q_2 + \delta_3 \Delta p_{female,gjc} q_3 + \Delta \varepsilon_{female,gjc} \qquad (6)$$

I then use an F-test to test that $\delta_1 = \delta_2 = \delta_3 = 0$

3.2 Weighting of data

This analysis uses several specifications, with some consisting of results from several key stages. This raises two issues. First, since the dependent variable is created by taking a mean of pupils' test scores, simply using this score unweighted would lead to a mis-specification of the model, as large schools would necessarily have the same weight in the model as small schools. Thus, the first part of the weighting is the number of pupils used to create this average score. The second issue is raised when I pool multiple key stages in the analysis; as for instance, there are only 3 observations of GCSE results, whilst there are 8 years of key stage 2 results. Since I take first differences, there is one fewer observation in the OLS specification, and so, I consider the number of cohorts less one. Thus, the second part of the weighting is to divide the weights by the number of cohorts, less one, that are observed for each key stage assessment. Furthermore, this only gives the weight required for each individual year, rather than for the change between years, so in order to deal with this, I take the average of the weightings for consecutive years.

i.e. The weight is calculated thus:

$$W_{male,gjc} = \frac{(N_{male,gjc} + N_{male,gjc-1})/2}{C_g} \qquad (7)$$

Where N is the number of male (female) pupils in the school and C is the number of cohorts observed at level g.

4. Data

There is a quasi market in English schools, where parents make a preference of which school they would like their child to attend. If the school is over-subscribed, then the school, or local authority, uses admissions criteria, which may be related to the distance from the family home to the school, whether the child has any siblings at the school, or in some schools their attendance at church or equivalent schools.

In England, it is compulsory for children to be in education up until age 16, and during this time they are taught in key stages as set out in the national curriculum. At the end of each key stage, the pupils are assessed by either teacher assessment or in nationally administered tests. These key stages are administered at age 7 (key stage 1), age 11 (key stage 2), age 14 (key stage 3) and 16 (key stage 4). During their school careers, pupils are taught in different types of schools, depending on their age. Whilst there are several different models of schooling in England, the most common is that pupils attend infant schools from age 4 to age 7, junior schools from age 7 to age 11, and secondary school from age 11 onwards. Often, infant and junior schools are combined to make one single primary school.

Pupils are examined in reading, writing and mathematics at KS1, English mathematics and science at KS2 and KS3, and in multiple subjects at GCSE. Pupils' achievement at KS1, KS2 and KS3 is measured in national levels. In each subject, the national curriculum is separated into strands which assess various skills within the subject, and each level is associated with a certain skill level that needs to be achieved. Levels can be achieved between 1 and 8, with a further grade only available for exceptional performance. These are converted into national curriculum scores.

Within GCSE, results are presented using the range of A* to G, with a U for a fail, A* indicating the highest grade and G the lowest. Whilst the GCSE grades are measured in a different way from the key stage levels, in order to quantify the results, I consider a comparable linear scale to that used in the key stages

I use data from the National Pupil Database (NPD) and the Pupil Level Annual School Census (PLASC), containing data on all pupils in state funded education in England. Pupil level characteristics that are collected include the pupils' age within the year, their gender, ethnic group, their exclusion status and a measure of low income with the free school meals (FSM) indicator. There are also school level characteristics such as the type of school, number of full time teachers within the school, whether there are pupils present who are boarders, etc.

The National Pupil Database (NPD) gives results of pupils in the key stage assessments. The structure of the available data is shown in Figure 1, with results available for pupils who sat key stage 1 (KS1) between 1998 and 2004, key stage 2 (KS2) between 1996 and 2004, key stage 3 (KS3) between 1998 and 2004 and GCSE between 2002 and 2004. The pupil-level data contained in PLASC, however, can only be linked to pupils who were in full time education when PLASC was initiated in 2002. Thus, the pupils who sat Key stage 2 in 1996, for example, have no PLASC data.

When conducting the data analysis, it was not possible to observe any pupils through the entire assessment process in schools with the data in PLASC[4], but I can observe the cohorts who sat key stage 2 in the three years, 1997, '98 and '99 at both key stage 3 and key stage 4.

In this paper, I wish to examine the effect of a change in the gender of the peer group on outcomes. However, within the data, I do not have any data on which pupils are taught in classrooms together, only which school they are taught in. In order to consider classroom level interactions, I take advantage of a property of infant and junior schools. In infant schools since the start of the 2001/2002 academic year there has been a legal requirement that there should not be more than 30 pupils to a qualified teacher (the Education (Infant Class Sizes) Regulations 1998[5]). Effectively, this means that the maximum class size in infant's school is 30. However, according to statistics from the Department for Children, Schools and Families[6] in 2009, 310 classes out of 53,160, or 0.6% of all key stage 1 classes, are taught in unlawful classes of 31 or more, whilst 580 classes (or 1.1% of all key stage 1 classes) were in lawfully allowed key stage 1 classes with more than 30 pupils. These lawfully allowed over-size classes are approved due to children being allowed entrance following appeals, being moved into a school's area after the start of the school-year or when the LA has placed a child with special needs into a school.

If infant schools follow this legislation, then we would expect to see a saw-tooth distribution of school cohort sizes. That is, we would expect to see peaks at sizes of 30, 60, 90 etc. This would indicate schools being filled until they reach the maximum size of 30 for a classroom, and if they have additional capacity, then starting a new class. Figure 2 shows the distribution

[4] This paper uses an early release of the NPD that covers up until 2004.
[5] See http://www.opsi.gov.uk/si/si1998/19981973.htm for more details. Accessed 10/8/2009
[6] See http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000843/SFR08_2009_ClassSizeCommentary.pdf for more details. Accessed 10/8/2009

of school sizes at key stage 1. It is clear that there is clustering at and below schools of size 30 and 60 at key stage 1, indicating that the schools are filling up the available spaces, and then stopping admissions. Figure 3 examines the distribution of school sizes at key stage 1 before the introduction of the legal limit of 30 pupils to a class in 2002, and figure 4 examines the data after the introduction. Whilst there is a more pronounced fall in the number of schools with 31 pupils compared to 30, post-2002 there is still a significant drop. Thus, at key stage 1, it seems a valid strategy to consider schools with 30 pupils as being schools which primarily teach all of their pupils within the school-year in one class.

At other levels, there is no legal maximum class size. However, figure 5 shows the distribution of school sizes at key stage 2. There is a similar distribution as that seen in key stage 1, but again with less pronounced falls after school sizes of 30 and 60. However, this evidence is sufficient to make the assumption that schools with 30 or fewer pupils consist of one class. Figure 6 shows the distribution of school sizes for key stage 3 and figure 7 shows the distribution of school sizes for key stage 4. It is clear that in secondary schools, no such strategy is available to us, as the size of schools is much larger.

Science at key stage 4 needs to be treated carefully. Not all pupils are assessed in the same way for science. There are three possible structures that are examined for science; one single award, covering all of physics, chemistry and biology, a dual award, which gives the students two identical grades, or up to three separate sciences. Thus, a student may receive 1, 2 or 3 grades at key stage 4 science. As such, to create a comparison across pupils, I consider the mean of their science scores.

4.1 Sample Selection

In order to control for endogeneities caused by selection, I only consider here schools in non-selective local authorities[7] (LAs); that is, where fewer than 10% of the pupils are selected by ability. For the purposes of this analysis, I consider a local authority that performs selection to be one where over ten percent of the pupils are in schools that select pupils according to ability. Whilst the non-selective schools in the selective local authorities do not select directly, due to the fact that there are schools in the same catchment area that have the opportunity to

[7] As defined in Atkinson et al (2006)

select pupils based on ability, the non-selective schools are left with a non-random selection of pupils. Furthermore, I only include community, voluntary aided, voluntary controlled, foundation, and city technology college schools. In addition, any school that has records of having boarding pupils is dropped as well.

It is apparent that some of the schools appear to have vastly different numbers of pupils from one year to the next. In order to prevent these outlying schools from adversely affecting the results, I only consider schools that lie within the 1^{st} to the 99^{th} percentiles of cross cohort changes in school sizes. That is, schools which have an improbably large change in size from one year to the next are removed from the sample. In real terms, at key stage 1, I only consider schools that fall by a maximum of 20 pupils from one year to the next and rises by 18, at key stage 2, a maximum fall and rise of 21, at key stage 3, a maximum fall of 43 or a rise of 52 and at key stage 4, a maximum fall of 34 and a rise of 54.

Further, some schools have very large (or very small) proportions of girls in the school. In order to remove the possibility that some of these schools have some sort of endogenous selection policy based on gender, schools that lie outside of the 1^{st} to 99^{th} percentile of the gender mix (after single sex schools are dropped) are also dropped. This leads to a range of the proportion of pupils that are female between 16.66% and 80%.

Finally, in order to have a consistent sample across the time series, only schools that appear in all of the observations are included. Thus, any school that closed, (or opened or failed to report results) during the time-period of the data is omitted. Table 1 shows the total number of schools available in the data, and the number of schools that remain once I have dropped observations as described above.

The raw data is presented in terms of national curriculum levels achieved by the pupils in the specific key stage, which should be comparable across years. In order to make the results easily comparable across key stages, the raw results are standardised by subject and level to a mean zero and standard deviation of one.

4.2. Summary statistics

Table 3 shows summary statistics for the entire sample, for secondary schools and for each individual key stage, whilst Table 4 shows summary statistics for small primary schools.

The scores from English, mathematics and science key stage assessments are presented in a weighted form, as described above. The proportion of girls within the cohort and the size of schools are weighted slightly differently, with the number of cohorts observed at each level used as the weighting. Whilst this does not affect the statistics within key stages, when they are pooled it does place more weight on the larger secondary schools. Science appears to have a lower sample size in the pooled specification simply because science is not assessed at key stage 1, whilst English and mathematics are.

Since all of the individual key stage results are based on means of normalised results centred at 0 with standard deviation 1, it is possible to compare the mean scores between key stages. Looking at the pooled data, it can be seen that on average girls perform much better than boys in English, but in mathematics and science, there is little or no difference, with boys initially holding an advantage, although girls overtake them by the time they reach key stage 4.

At key stage 1, girls have a significant advantage in English, whilst there is little difference between the genders in mathematics. At key stage 2, there is still a significant advantage for girls in English, whilst in mathematics and science, the boys hold a small advantage. At key stage 3, the gap between the genders is increased in English, and boys still hold a very slight advantage in mathematics and science. However, this changes slightly at key stage 4, with girls maintaining a large significant advantage in English, but taking a small lead in mathematics and science as well.

The gender mix in the schools remains constant at approximately 48% to 49% female throughout, with cohort sizes within school of approximately 40 at key stages 1 and 2 and approximately 180 at key stages 3 and 4. This may make inferences at a school level much harder at the secondary level due to the fact that whilst there may be a larger proportion of female pupils in one school than another, individual pupils may not feel the effect of this due to a lack of within school interaction. That is, at a cohort level there could be a large proportion of girls, but this may not permeate down to the classroom level, whether due to ability setting or some other mechanism.

Table 4 shows the summary statistics for key stage 1 and key stage 2 in small primary schools. As with Table 3, the proportion of the cohort that is female is approximately 0.49. By placing the restriction on considering only single classroom primary schools, I keep 23 percent of the schools.

5 Results

In looking at the results, I start by looking at a specification that includes the full sample of schools, and all of the available key stages, followed by tests of linearity of the specification. This is then followed by specifications solely including primary and secondary schools, then results by the individual key stages. I then examine effects in small and large primary schools to try to examine the effect of the direct peer influence, and then examine the effects within key stage within small schools. I follow this up by examining the robustness of the results by comparing them with those from a subset of the sample that only contains schools which appear to have completely random changes in the gender make-up from year to year. Finally, I repeat the specifications using a value added model to examine the effects of a change in the gender make-up of the peer group on the value added from one key stage to another.

5.1 Results in all schools

Table 5 shows regression results for all schools in English, mathematics and science. The initial specification includes all schools and levels, and I estimate equation (5) using analytic weightings, as specified in equation (7):

$$\Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \Delta \varepsilon_{female,gjc} \qquad (5)$$

where p is the proportion of pupils in the cohort within the school that is female and X includes year dummies and dummies for the key stage level.

In English, there is a significant negative effect for male pupils of having a more female peer group, whilst for mathematics and science; girls boys experience a significant positive effect of having a more female peer group. Girls appear to be unaffected by having a more female peer group in English. If one considers the effect of the proportion of girls increasing by 10%, then these raw effects would lead to a fall in average English scores for boys by approximately 0.016 standard deviations. For girls, this 10 percentage point increase in the proportion of the peer group that is female would lead to an increase in average mathematics score by 0.007 standard deviations and in science by 0.010 standard deviations. Furthermore, when translated into the effect of a more able peer group, I obtain coefficients that imply that in English, a 1 point increase in the average ability of the peer group is associated with a 0.23 point drop in the individual boys' outcomes, whilst for mathematics, a similar increase in peer ability is associated with a 1.66 point drop in girls average outcomes respectively[8], whilst for science, a 1 point increase in the average ability of the peer group is associated with a 3.4 point decrease in individual girls' outcomes.

In order to check that the model is valid, it is necessary to examine the linearity of the estimates for the coefficient on the proportion of pupils that is female. Figure 8 shows adjusted variable plots for the pooled regressions in English. In the graphs, the x axis represents the 50 quantiles of the proportion of pupils that is female within the cohort in the school and the y axis represents the mean change in average outcomes that can be attributed to just the change in female proportion for each quantile. The fitted line is the fitted regression line. For boys, the adjusted variable plot shows a negative slope, but for girls no real trend is observed. **Figure 9** and Figure 10 show the adjusted variable plots for mathematics and science respectively. These suggest a positive relationship between the proportion of girls within the classroom and both boys and girls outcomes in mathematics and science.

Table 6 reports the results of regressions for equation (6) using the pooled specification above, including terms interacting the proportion of pupils that is female with the quartile that it is in. Using an F-test, in all of the subjects for all of the pupils, I do not reject the null that the coefficients on all of these interaction terms are equal, and equal to zero, so I do not reject the null of linearity.

[8] Methodology for calculating this figure, as used by Hoxby (2000) is detailed in the appendix

Table 7 shows results of regressions within primary schools, secondary schools and by key stage. Beginning with primary schools, there is a strong significant negative effect on increasing the proportion of pupils within the cohort that is females on male pupils in English. This same effect is seen in both key stage 1 and key stage 2 results, although the magnitude of the coefficient at key stage 1 is much larger than that at key stage 2. For mathematics and science, both male and female pupils see a significant and positive effect of a more female peer group in primary schools. Within secondary schools as a whole and at the finer level of key stages 3 and 4, the only significant effects of a more female peer group are a strongly negative effect for males in English. In terms of the size of effect, at key stage 1, a 10 percent point increase in the proportion of pupil who are female has the following effects; a 0.018 standard deviation fall in boys' average outcomes in English, a 0.011 standard deviation increase in boys' and a 0.007 standard deviation increase in girls' average outcomes in mathematics. For key stage 2, the proportion of girls in the class has no impact on boys' outcomes in English whilst girls' average outcomes are increased by 0.009 standard deviations in English with a 10 percentage point increase in the proportion of the peer group that is female. In mathematics, both boys' and girls' average outcomes are raised by 0.006 and 0.007 standard deviations respectively, whilst for science boys' and girls' outcomes are raised by 0.012 and 0.009 standard deviations respectively. For key stages 3 and 4, a 10 percentage point increase in the proportion of pupils who are female is associated with a 0.033 and 0.028 standard deviation fall in boys' average outcomes in English[9].

Table 8 shows the results translated into the effect of a change in the ability of the peer group based on the change in the gender make-up of the peer group. As with the results presented in Table 5, I find large, negative, effects of an exogenous increase in the peer group ability in mathematics and science at key stages 1 and 2. These effects are of a much larger magnitude than would be credibly expected, and are in the opposite direction to those expected. As with Hoxby (2000), and as suggested by Lavy and Schlosser (2007), the magnitude and direction of these translated effects suggest that something other than merely an ability spillover is occurring. As such, I do not consider these translated effects any further, and merely concentrate on the effect of a 'more female' peer group on outcomes.

[9] I have not considered here any result that is insignificant at the 10 percent significance level.

5.2 Small and large primary schools

Table 9 shows results from the subset of schools that are either defined as small or large schools. The small school definition is a school that does not go over the limit of 30 pupils within the cohort in any year observed in the data, indicating a very high probability that there is only one class within the cohort, and the large school is any school that has more than 30 pupils in all of the observed cohorts, indicating multiple classes. The results for the large primary schools are not significantly different from those for primary schools as a whole. However, within the small primary schools, there is a much larger negative effect on boys in English of a more-female peer group. However, much of the larger magnitude can be explained by the much larger coefficient within key stage 1 scores in small primary schools, which is approximately two and a half times as large as the coefficient for key stage 2. This difference may be explained as results at key stage 1 are generally more noisy than those at other key stages. The only other significant effect within small primary schools is a positive effect for girls in mathematics, which again is being driven by a large effect at key stage 1.

5.3 Robustness checks

It is possible that some schools have selection policies based on the gender of pupils, which could affect the results that are gained for the effects of a more female peer group on outcomes. In order to check that the results are not biased by unobserved selection policies, Table 10, Table 11 and Table 12 show comparisons between regressions with all of the schools included, for all of the specifications described above, and a subsample of schools which have apparently random changes in the gender make-up of cohorts. In general, the full sample results do not significantly differ from the random. In Table 10, for English, there are no major differences between the full sample and the apparently random sample. In mathematics, there is a small difference between the results in all levels and schools pooled for males, but this is generally an insignificant difference. Similarly, there is a difference of a reasonably large magnitude between the full sample and credibly random sample for boys in mathematics at key stages 3 and 4, but these are not significant differences. Similarly, there is a large difference between the random and the full sample for key stage 3 for males in science, but again, this is not significant.

Table 12 shows the comparison between the sub-sample of schools that have apparently random changes in gender make-up in the small and large primary schools. There is only one significant difference between the two sets of results, and that is for females in English at key stage 2. However, neither the result in the credibly random sample, nor the result in the full sample is significantly different from 0, so it does not affect my results.

5.4 Value added results

Table 14 shows the results of the estimation of equation (7), with the dependent variable as the average within cohort male (female) value added from one key stage to the next for pupils that stay within the same school, except between key stages 2 and 3, since almost all pupils are registered at a different school between these exams. Beginning by looking at the results for all schools and levels pooled, the only coefficients that are significantly different from zero are for females in mathematics and science. Examining the results at a finer level, it can be seen that this overall result is being driven by a large effect of a more female peer group on value added from key stage 2 to key stage 3, which also drives the large value added observed in the secondary schools for girls in mathematics and science.

In English, a more female peer group has a positive effect on boys and girls at key stage 2. However, comparing the regression results in Table 14 with those from Table 7 it can be seen that this may be due to pupils being disadvantaged at key stage 1 by having a more female peer group, and this disadvantage being reduced over time, with it actually becoming an advantage for girls. However, any advantage gained by girls from having a more female peer group from key stage 2 to key stage 3 seems to be eliminated between key stages 3 and 4, with a large significant negative effect on the proportion of pupils that is female. Finally, examining this value added measure in small primary schools, Table 15 shows the results and there is no observed significant effect of a more female peer group on this value added measure at key stage 2 in small schools

5.5 Validity of results.

The identification strategy utilised in this paper depends on year-on-year changes in the gender make-up of the cohort being random. However, having observed the gender make-up of the cohort, it is possible that a parent may decide to move their child to a different school based on the proportion of pupils who are female, which could result in non-random changes in the school gender mix. In order to check this possibility, I use a pupil level dataset and examine the gender make-up of a child's peers at age 7, and whether this is correlated with the decision to change school between key stage 1 and key stage 2. I consider a peer group at age 7 to be male-dominated if more than 60% of the peer group is male, and similarly a female dominated peer group to be one with 60% or more female. Some schools which enter pupils for both key stage 1 and key stage 2 examinations admit more pupils into year 3 by creating new classes. In order to remove these from the sample, I only consider schools where 80% or more of the pupils at key stage 2 were also in the school at key stage 1. For boys, the correlation with having a male dominated peer group at age 7 and being in the same school at age 11 is -0.0008 (P=0.5381), whilst for girls, the correlation is -0.0042 (P=0.0008). So for girls, there is a significant, but very small correlation between being in a male dominated cohort at age 7 and changing schools for key stage 2. For female dominated peer groups, for boys, the correlation between having as female dominated peer group and being in the same school at age 11 is 0.0064 (P=0.0000), whilst for girls it is 0.0053 (P=0.0000). So, again there is a significant, but very small correlation between being in a female-dominated peer group and staying in the same school. However, whilst there is a statistically significant correlation, the correlation is not quantitatively significant and since they are unlikely to have overly affected the results.

6 Conclusions

In this paper, I have examined the effect of seemingly exogenous changes in the gender make-up of a child's within-school peer group using year to year changes in the proportion of girls within the school as an explanatory variable for the outcomes at key stage 1, key stage 2, key stage 3 and GCSE.

The results show significant negative effects of a more female peer group on male pupils in English, robust across specifications, and a significant positive effect of a more female peer group in mathematics and science for both males and females in primary schools. When assessing the magnitude of these estimated effects, I compare the results with the effect of being older within an academic year. Proud (2010) shows that pupils born one month earlier within the school year perform between 0.01 and 0.05 standard deviations better than their younger peers. The results for English suggest that a 10 percentage point increase in the proportion of pupils who are female is associated with a between 0.01 and 0.03 standard deviation decrease in boys' scores, which is of the same magnitude of being 1 month younger within the school year. Boys and girls perform 0.01 standard deviations better in maths and science from a 10 percentage point increase in the proportion of pupils who are female, which is of the same magnitude of being 1 month older within the school year.

Lavy and Schlosser (2007) show that if their results looking at the effect of a more female peer group were driven by changes in the ability of the peer group due to this change in the gender mix, then, in contradiction with the established literature, an increase in the ability of the peer group would lead to a decrease in the pupils' outcomes. They argue that there is some other factor, such as behaviour, that affects the students' outcomes. They demonstrate that an increase in the proportion of girls leads to general increases in academic outcomes, and find that the presence of more female peers lowers classroom violence, whilst improving inter-student, and student-teacher relationships. However, this is not attributed to an individual improvement in behaviour, but rather a compositional effect. This would help to explain my results in mathematics and science, but not for male students in English.

The change in the gender make-up of the peer group could have an influence on the behaviour within the classroom. Younger and Warrington (1996) consider the interactions within the classrooms and the behaviour associated with boys and girls in the classroom. For boys there is an apparent stigma associated with working hard. Furthermore, there is also evidence that boys require more behavioural management than girls. According to the data in PLASC, 70.9 percent of children with statements of special educational needs are boys. This is further shown by the fact that 5 times as many boys are permanently excluded from schools than girls. However, these figures may be slightly misleading, as it has been conjectured that there has been an over-identification of special educational needs (SEN) in boys and a similar under-identification of SEN in girls. In addition, Francis (2000) concludes that boys tend to be louder and more demanding within the classroom, but rather than this directly hindering the boys' own outcomes, it may be having a detrimental effect on all of their classroom peers.

Whilst not being affected directly by their peers, the gender make-up of classrooms may lead to differential teaching methods within the classroom. Whilst teachers may believe that they do not use different methods with girls and boys, Younger et al (1999) find evidence that boys and girls are treated very differently in the classroom. Students claim that boys receive more negative attention than the girls, and there is evidence that teachers have a lower tolerance level to boys' behaviour than to girls, which can "lead to male disillusionment and a negative reaction to learning". (Younger et al 1999, 339) However, they also comment that there is little evidence in observed lessons that boys are given "more support than girls in the teacher-learning process" (Younger et al 1999, 339). Furthermore, Dee (2007) finds that girls taught in a classroom with a female teacher and boys taught with a male teacher tend to perform better than pupils with a teacher of the opposite gender, suggesting that female teachers may direct learning in a way that is more likely to benefit girls rather than boys. This, when combined with Macleod (2005) who comments that only 15.7% of all primary school teachers in England are male and half of 5 to 11 year olds have no contact with male teachers implies that girls are likely to benefit more in education due to the gender of teachers.

Considering the difference between the single classroom cohorts in primary schools with the full sample, there is a much larger magnitude negative effect within the single classroom case for boys in English, which tends to lead us towards the conclusion of more behavioural issues with boys, or possibly the impact of a more female orientated teaching method, leading to disadvantages for boys. Further, it appears that girls benefit from an environment more suitable for learning in mathematics if there are more girls in the classroom, whether through better behaviour or more directed teaching. This model has less noise in it than the larger schools, as I can observe directly the within classroom peer group. In the large schools[10], the negative effects for boys in English disappear, but apparent positive effects are seen in mathematics and science which are not seen in the small schools' case.

Since the methodology used here takes advantage of variation in the proportion of the cohort that is female, it is not possible to make definitive conclusions of the effect of educating pupils in single sex classrooms. However, the results obtained suggest that boys would benefit at all ages from being taught English in English schools with as small a proportion of girls as possible. In mathematics and science, the results shown here tend to imply that both boys and girls benefit from having more girls in the classroom. However, it is not possible to increase the proportion of girls for both boys and girls, implying that a mix of the genders is optimal in both mathematics and science.

[10] Schools with more than 30 pupils in the cohort.

References

Burgess, S., McConnell, B., Propper, C. and Wilson, D. (2004). Girls rock, boys roll: An analysis of the age 14-16 gender gap in English schools. *Scottish Journal of Political Economy* 51(2): 209-229.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. and York, R. L. (1966), *Equality of Educational Opportunity*, US Department of Health, Education & Welfare.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources* 42(3): 528-54.

Francis, B. (2000). *Boys, Girls and Achievement: Addressing the Classroom Issues*, Routledge Falmer, London.

Gorard, S., Rees, G. and Salisbury, J. (2001). Investigating the patterns of differential attainment of boys and girls at school. *British Educational Research Journal* 27(2): 125-139.

Hallinan, M. T. and Sorensen, A. B. (1987). Ability Grouping and Sex-Differences in Mathematics Achievement. *Sociology of Education* 60(2): 63-72.

Hansen, Z., Owan, H. and Pan, J. (2006), The Impact of Group Diversity on Performance and Knowledge Spillover - An Experiment in a College Classroom, *National Bureau of Economic Research,* Working Paper 12551

Hoxby, C. (2000), Peer Effects in the Classroom: Learning from Gender and Race Variation, *National Bureau of Economic Research,* Working Paper 7867

Hoxby, C. (2002). The Power of Peers. *Education Next* 2(2): 57-63.

Lavy, V. and Schlosser, A. (2007), Mechanisms and Impacts of Gender Peer Effects at School, *National Bureau of Economic Research,* Working Paper 13292

Machin, S. and McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy* 21(3): 357-372.

Macleod, D. (2005). "Drive to recruit more male primary teachers." *Guardian Unlimited* Retrieved 21/08/09, from http://education.guardian.co.uk/primaryeducation/story/0,,1591234,00.html.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60(3): 531-42.

Proud, S. (2010), Evidence of Peer Effects in English Schools, *University of Bristol,* mimeo

Shibley Hyde, J., Fennema, E. and Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin* 107(2): 139-155.

Summers, A. A. and Wolfe, B. L. (1977). Do Schools Make a Difference? *American Economic Review* 67(4): 639-52.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review* 95(2): 199-203.

Winkler, D. R. (1975). Educational Achievement and School Peer Group Composition. *Journal of Human Resources* 10(2): 189-204.

Younger, M. and Warrington, M. (1996). Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school. *British Journal of Sociology of Education* 17(3): 299-313.

Younger, M., Warrington, M. and Williams, J. (1999). The gender gap and classroom interactions: reality and rhetoric. *British Journal of Sociology of Education* 20(3): 325-341.

A.1 Estimating the effect of a change in the ability of the peer group.

Measuring the effect of peer ability.

The three mechanisms for peer effects are discussed above. We wish to examine the apparent effect of the peer group ability on outcomes. In order to do this, we need to begin with the assumption that all of the gender peer effects are due to the change in ability of the peer group.

In order to examine this, we first need to establish the average overall outcome which is not affected by our peer ability measure, which we can calculate by using the estimated coefficient from the proportion of pupils that is female.

$$\hat{\mu}_{female,g} = weightedmean(A_{female,gjc} - \hat{\gamma}_{female,g}\, p_{female,gjc})$$

Now, if females on average score 1 point higher than males, then an increase in the proportion of pupils that is female of 10% would lead to an increase in peer ability of 0.1 points.

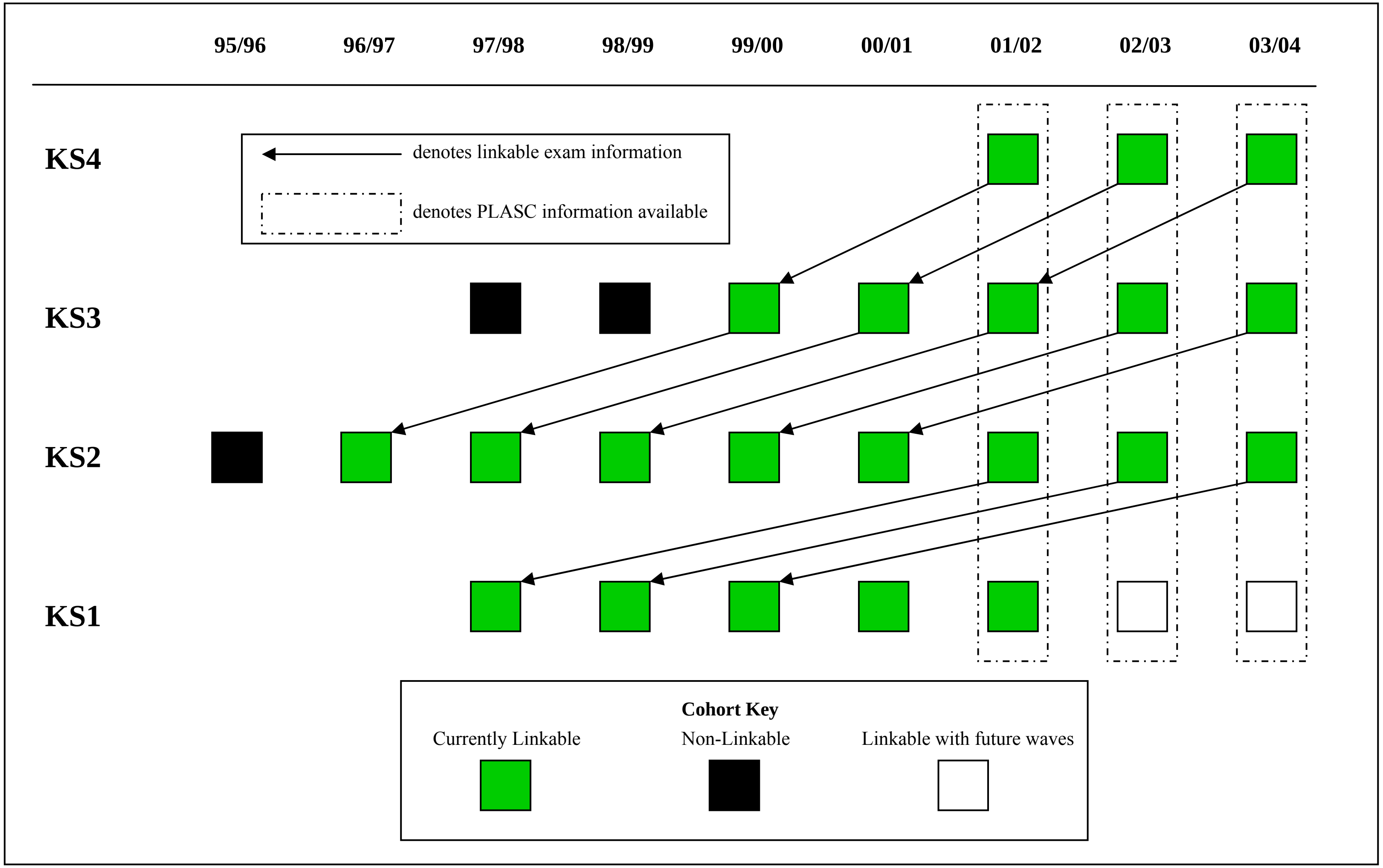
Thus, we can calculate the percentage change in the proportion of pupils that is female required to produce a 1 point increase in the peer group ability.

$$change = \frac{1}{\hat{\mu}_{female,g} - \hat{\mu}_{male,g}}$$

Thus, in order to calculate the effect of a change in peer group ability of 1 point, we multiply the change by the coefficient on the proportion of pupils that is female.[11]

[11] This methodology is the same as that employed in Hoxby (2000)

Figure 1 **The Structure of the cohorts available in the NPD and PLASC**[12]



[12] From http://www.bris.ac.uk/depts/CMPO/PLUG/userguide/cohorts.pdf

Figure 2 **Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils.**

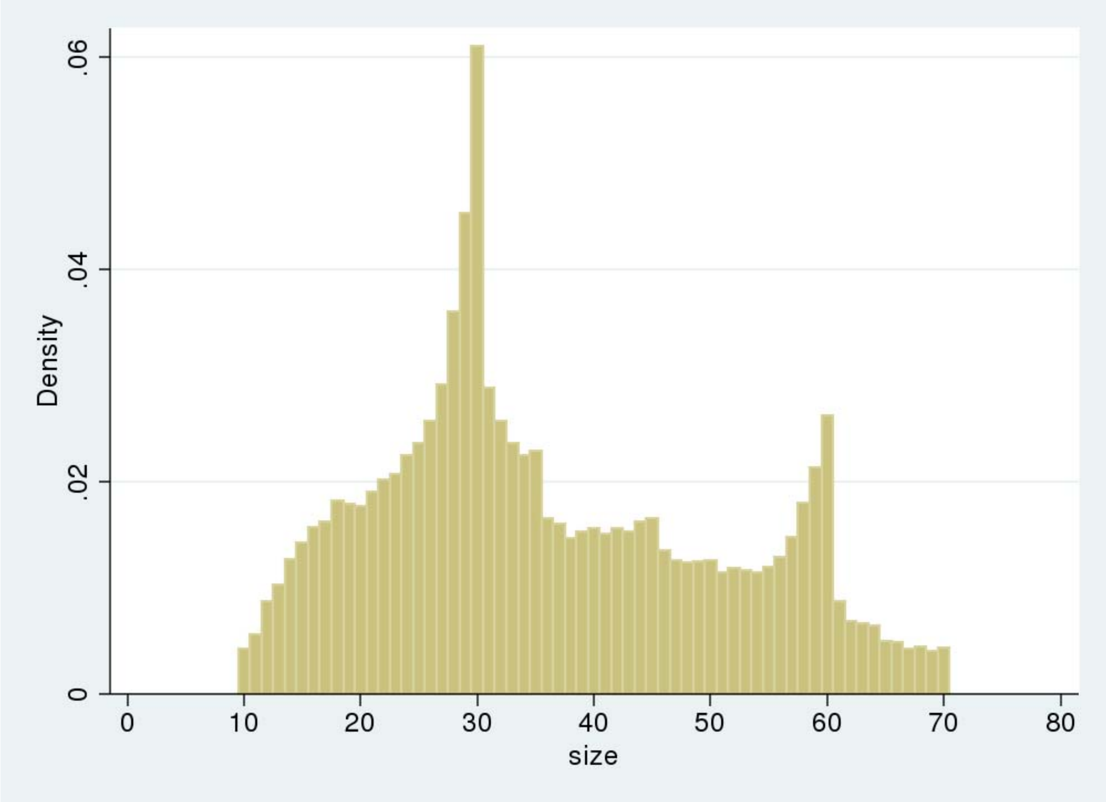


Figure 3 **Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils. Pre 2002**

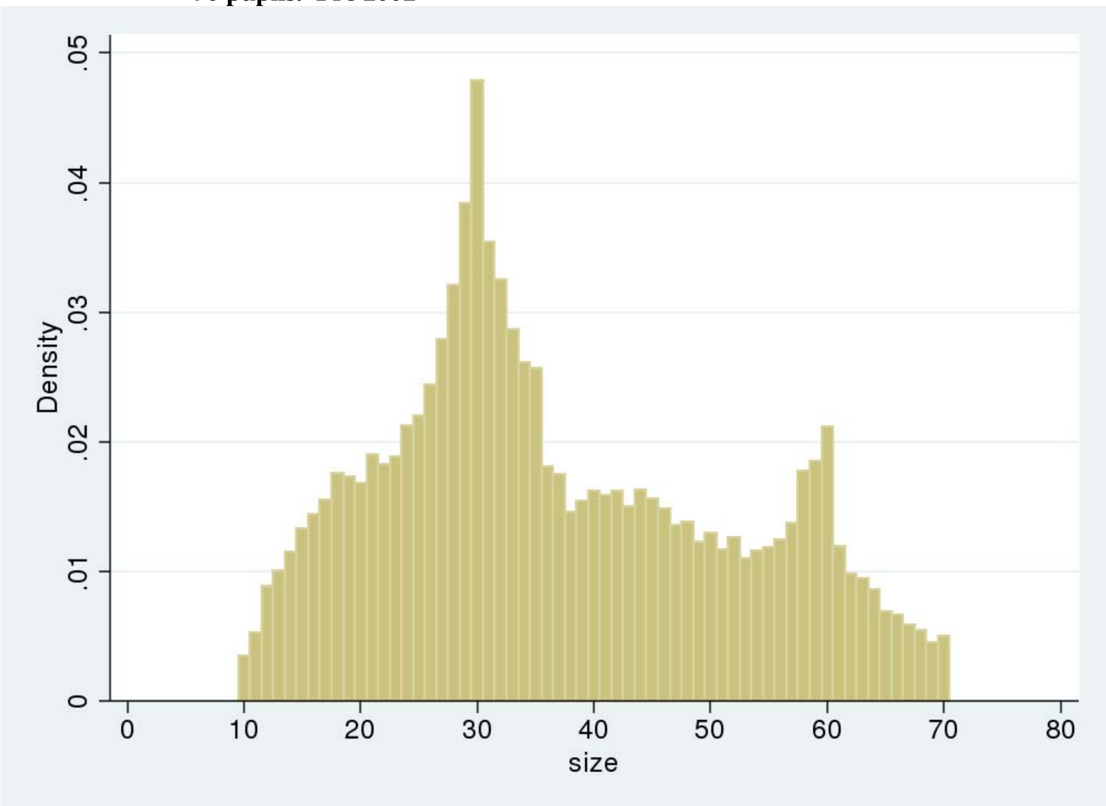


Figure 4 **Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils. Post 2002**

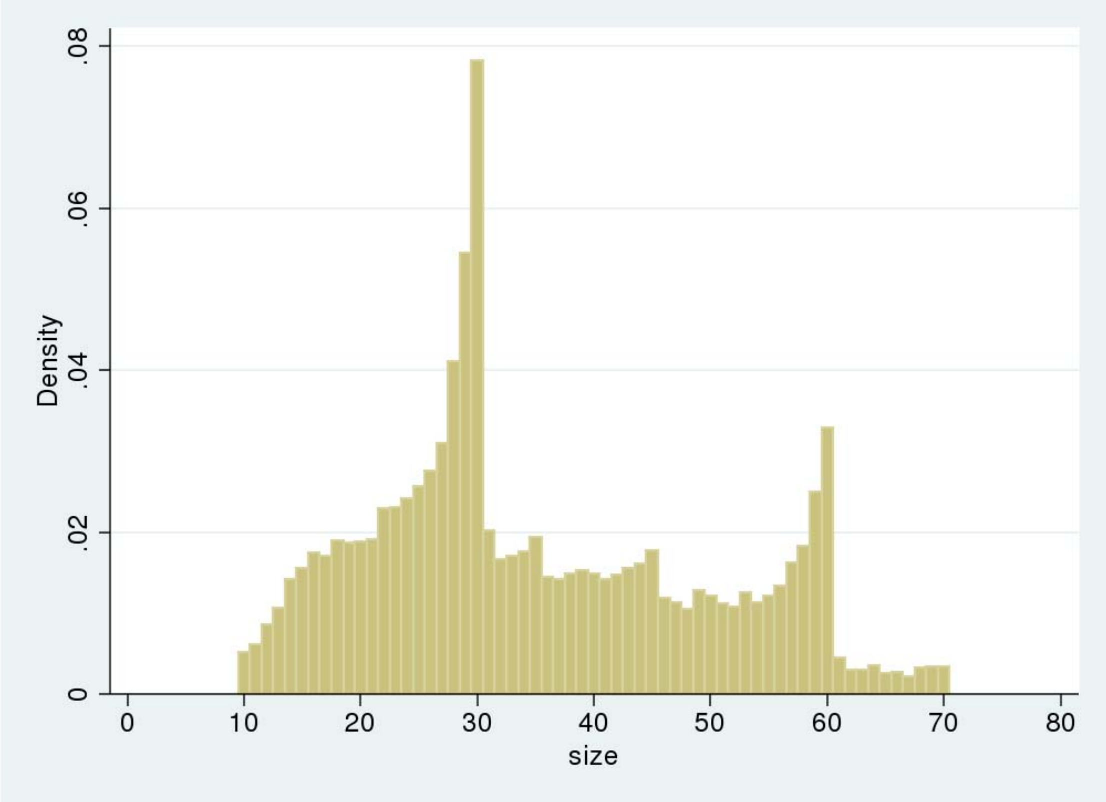


Figure 5 **Distribution of school sizes at Key Stage 2 focussing on schools with fewer than 70 pupils**

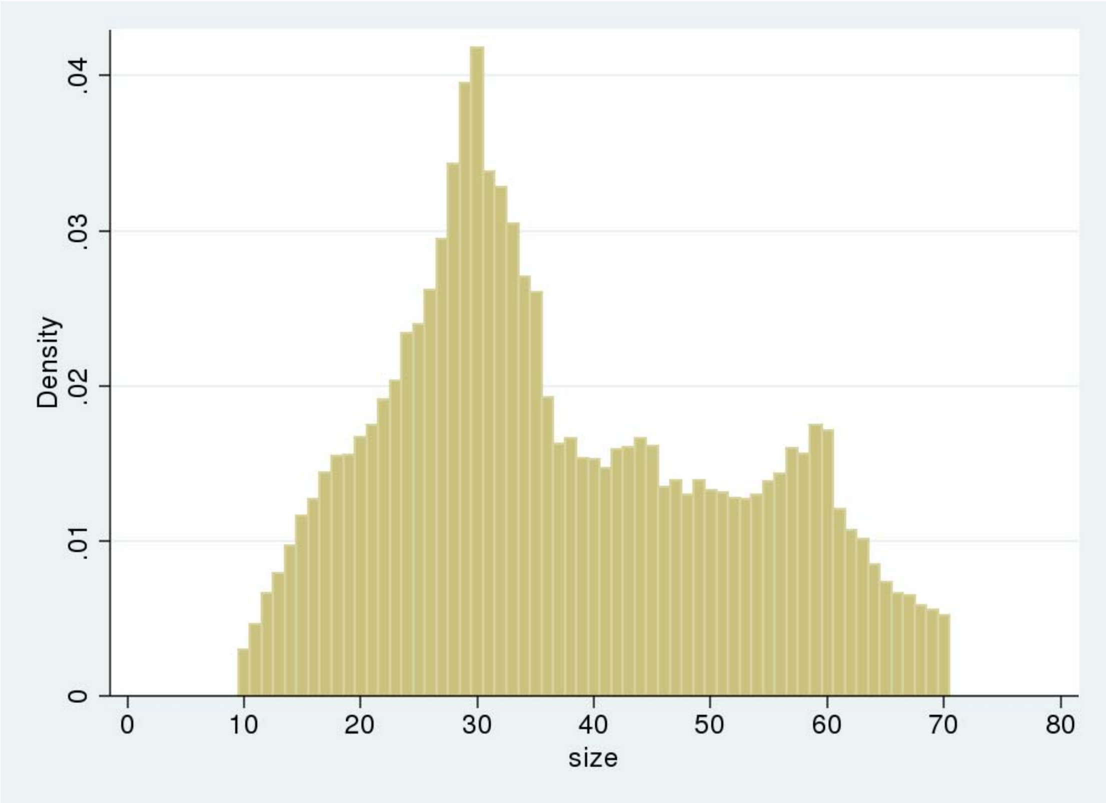


Figure 6 Distribution of school sizes at Key Stage 3

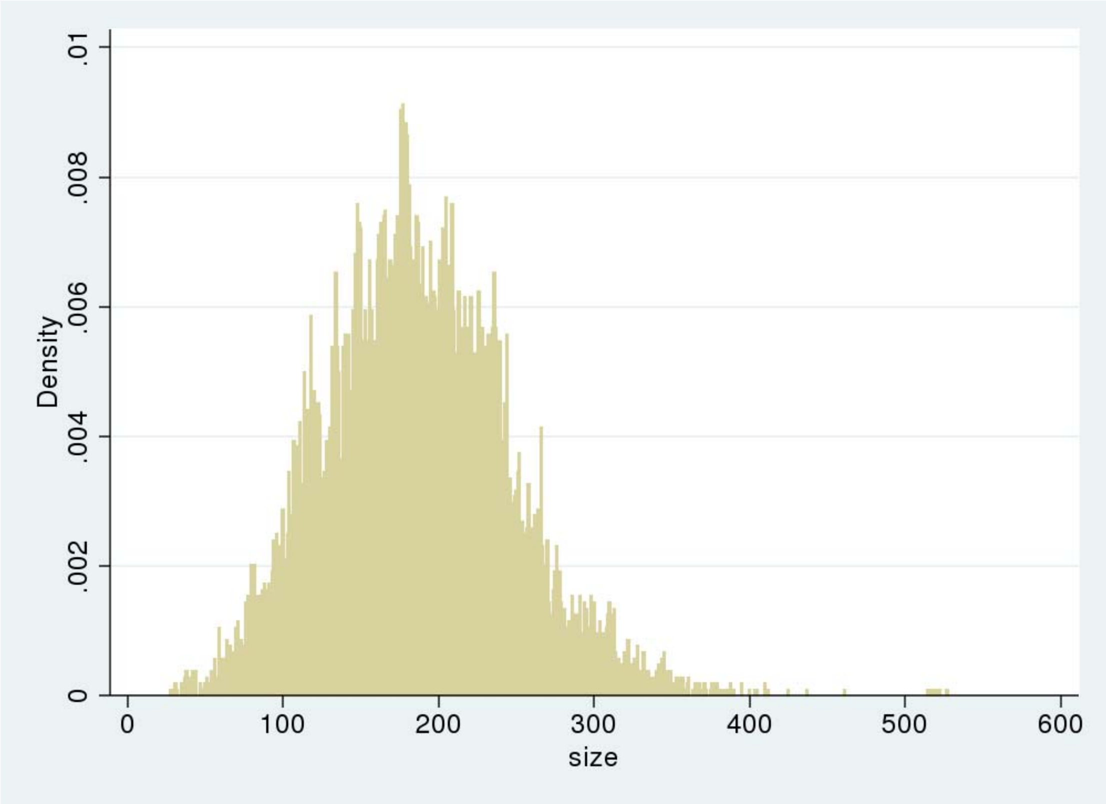


Figure 7 Distribution of school sizes at Key Stage 4

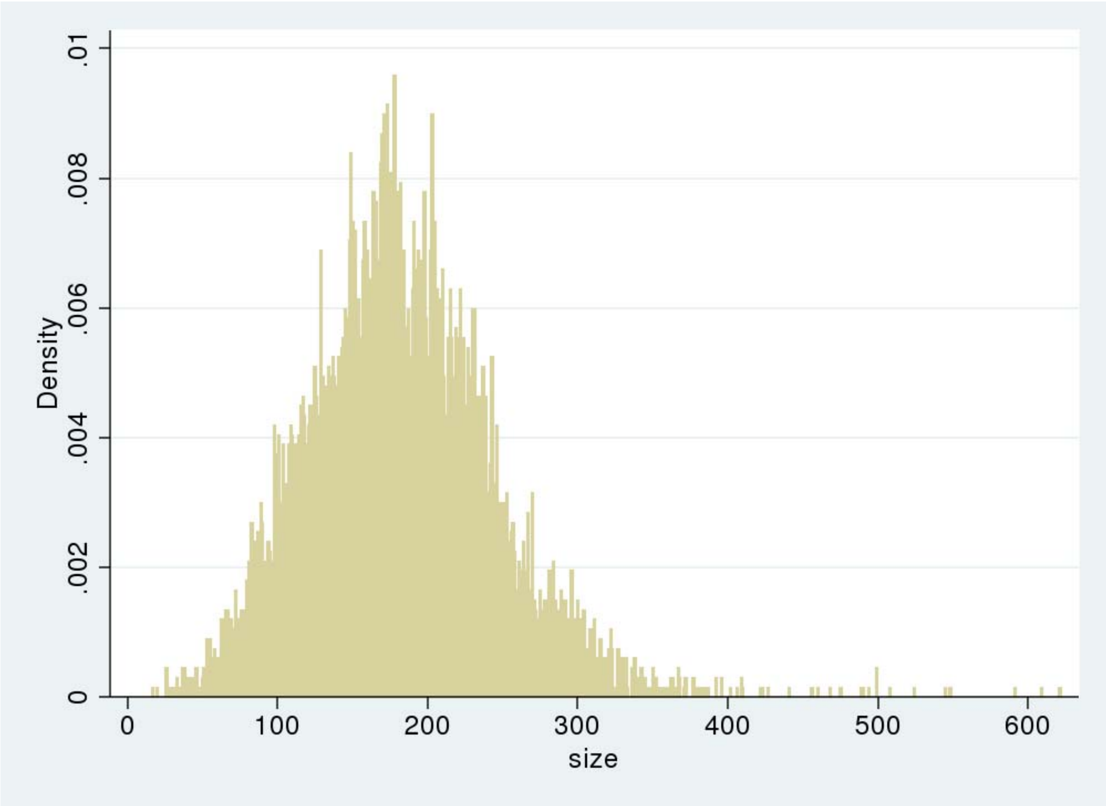
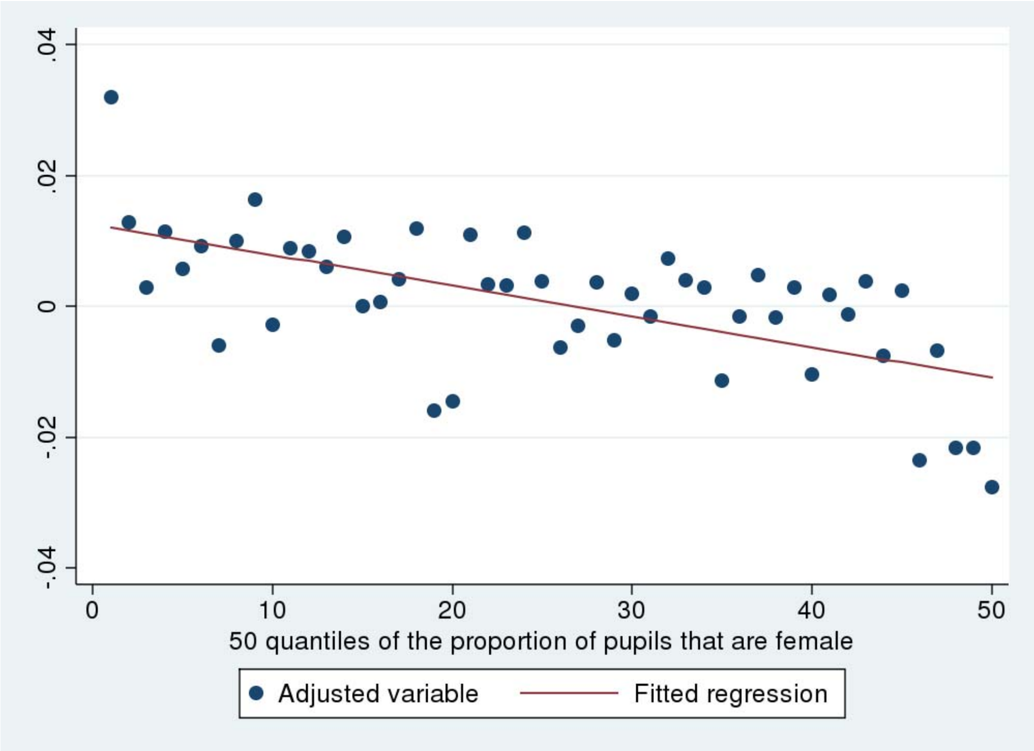


Figure 8 Adjusted variable plots of the pooled regressions for English

Boys



Girls

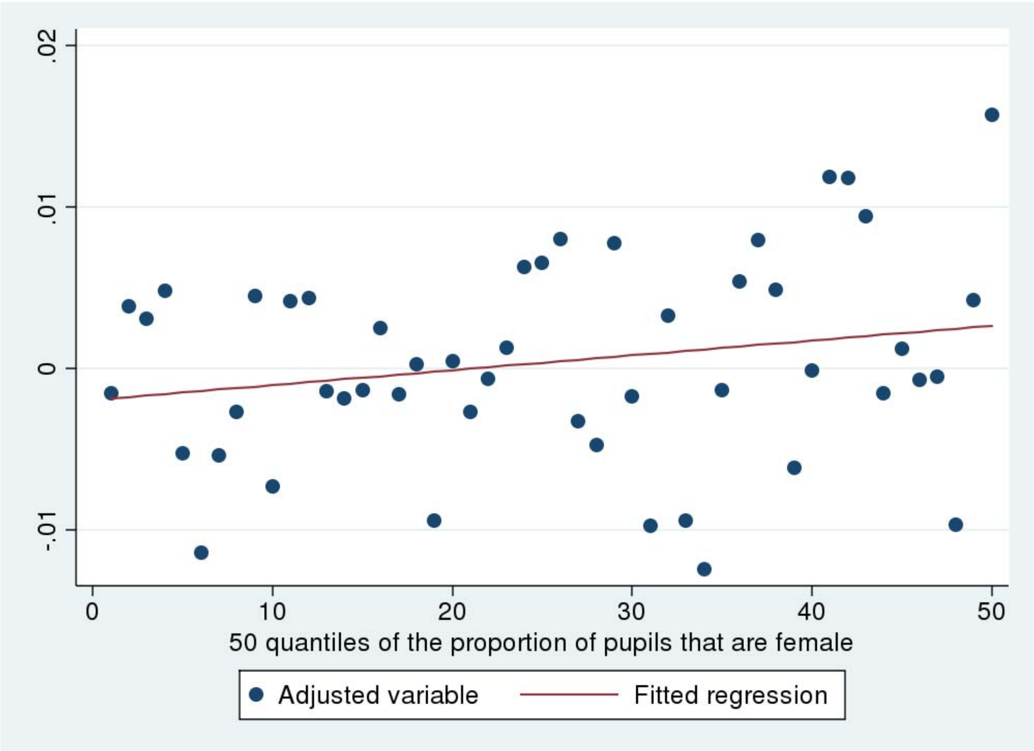
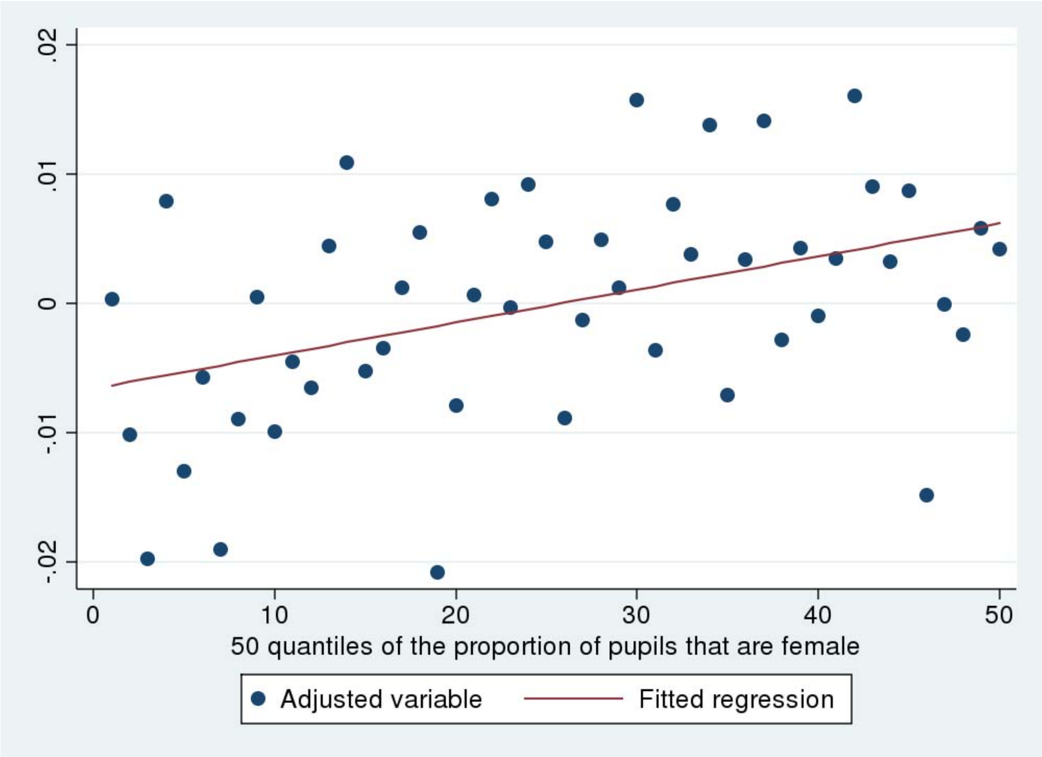


Figure 9 Adjusted variable plots of the pooled regressions for mathematics

Boys



Girls

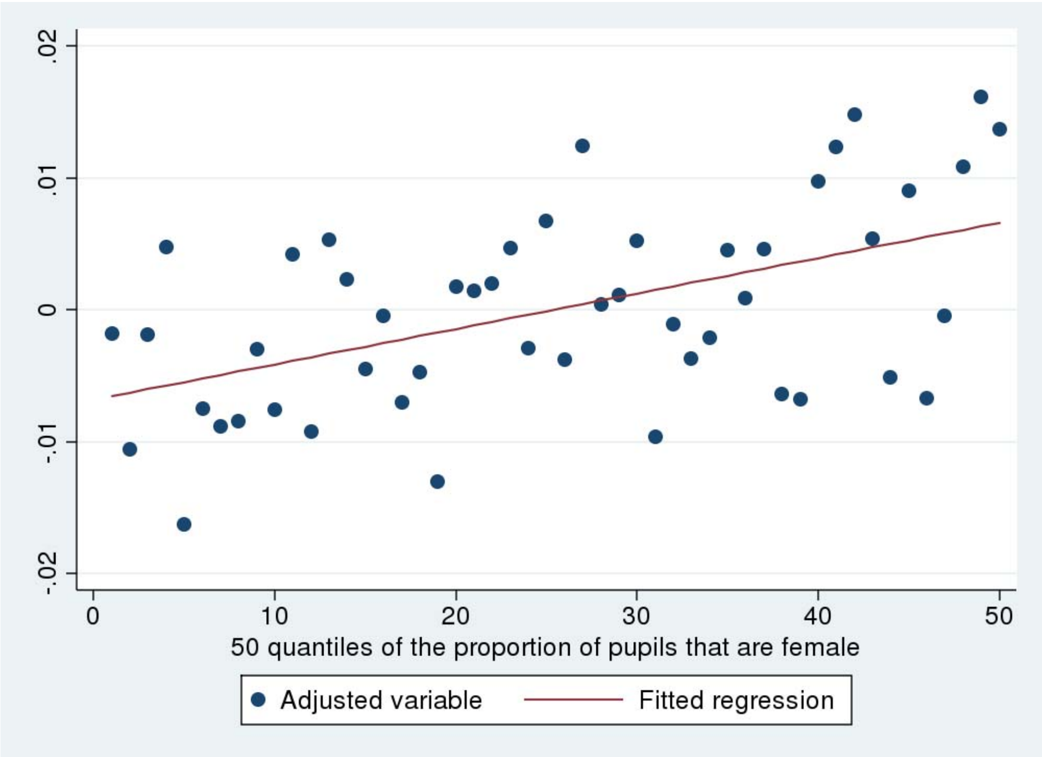
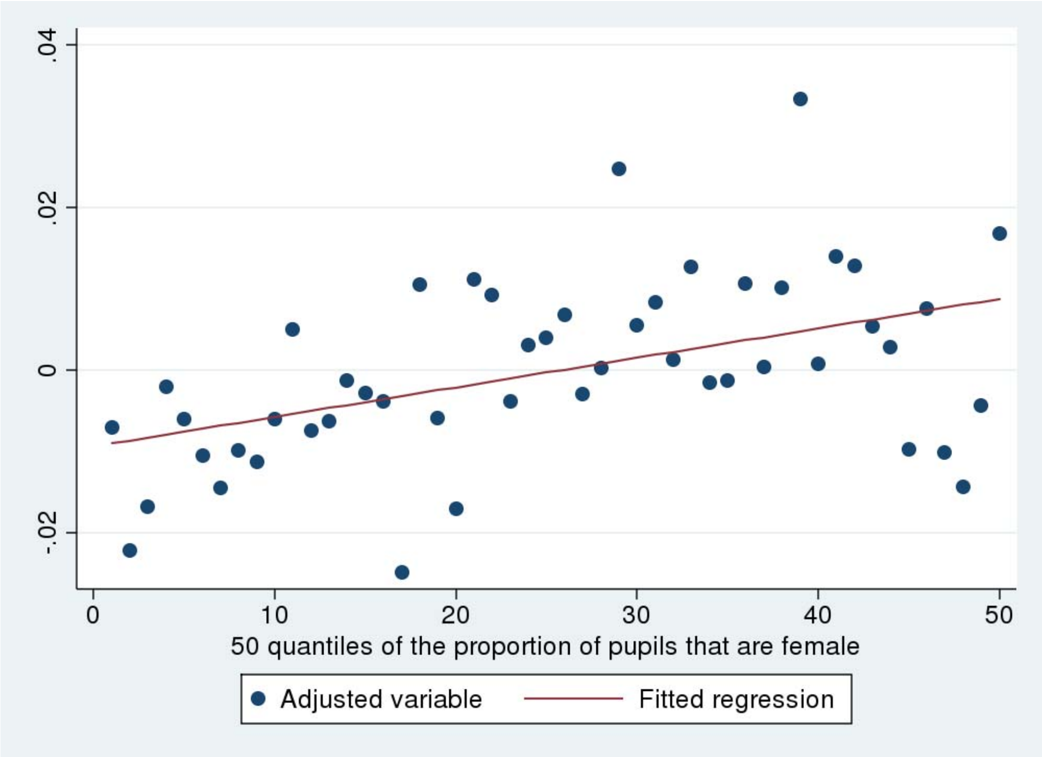


Figure 10 Adjusted variable plots of the pooled regressions for science

Boys



Girls

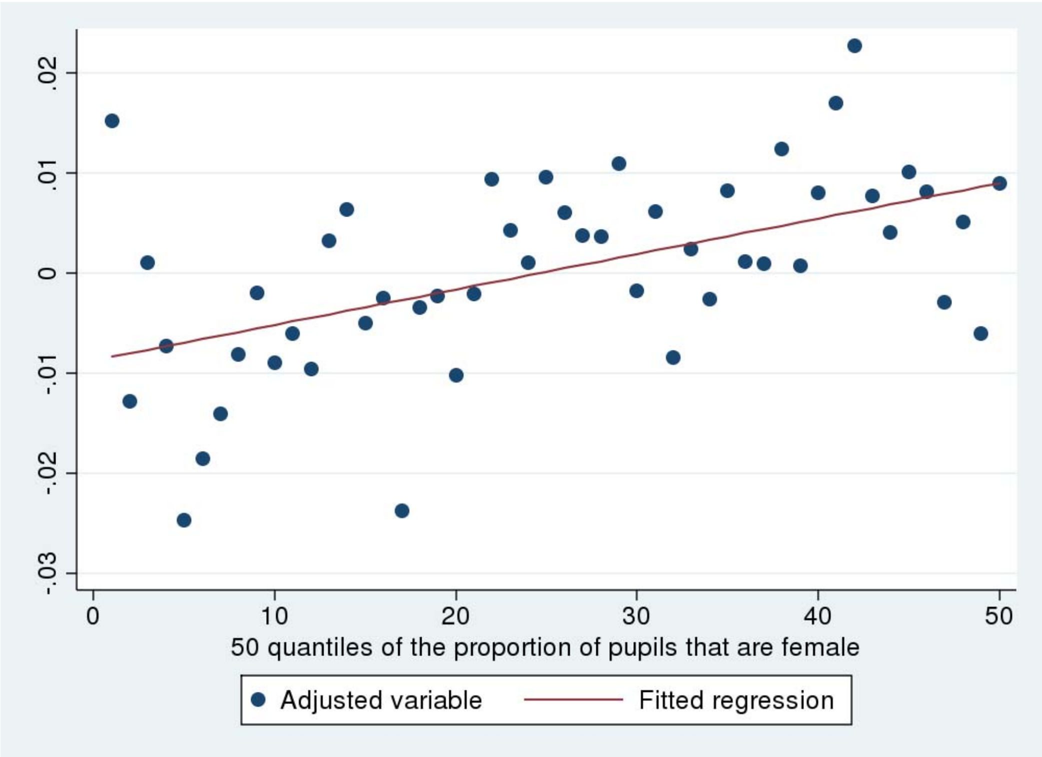


Table 1 Number of schools in the dataset by level before and after schools are omitted.

Level	1996	1997	1998	1999	2000	2001	2002	2003	2004
Key Stage 1									
Full Sample			16964	17093	17397	17150	16974	16887	16783
Sample after observations dropped			10669	10669	10669	10669	10669	10669	10669
Key Stage 2									
Full Sample	16013	16552	16782	16730	16732	16768	16514	16800	16461
Sample after observations dropped		9499	9499	9499	9499	9499	9499	9499	9499
Key Stage 3									
Full Sample			4529	4507	4600	4583	4447	4628	4493
Sample after observations dropped					2227	2227	2227	2227	2227
Key Stage 4									
Full Sample							3503	3481	3483
Sample after observations dropped							2335	2335	2335

Table 2 Proportion of pupils that stay at the same school between key stages

	Key Stage 1 – 2	Key Stage 2 – 3	Key Stage 3 – 4
Total	911,470 100%	1,359,182 100%	1,066,189 100%
Pupils at same school	600,814 65.9%	5,380 0.4%	1,032,461 96.8%
Pupils at different school	310,656 34.1%	1,353,802 99.6%	33,728 3.2%

Table 3 Summary statistics

	Mean score for males in English	Mean scores for females in English	Mean scores for males in mathematics	Mean scores for females in mathematics	Mean scores for Males in science	Mean scores for females in Science	Proportion of the cohort that is female	Size of cohort within school
Pooled Specification								
Mean	-0.174	0.182	0.006	-0.006	0.002	-0.002	0.488	112.566
Standard Deviation	(0.419)	(0.400)	(0.408)	(0.404)	(0.431)	(0.439)	(0.072)	(84.753)
Observations	160604	160604	160604	160604	89344	89344	160604	160604
Key Stage 1								
Mean	-0.156	0.164	0.005	-0.004			0.489	39.350
Standard Deviation	(0.432)	(0.401)	(0.434)	(0.408)			(0.090)	(19.030)
Observations	71260	71260	71260	71260			71260	71260
Key Stage 2								
Mean	-0.150	0.154	0.032	-0.033	0.007	-0.007	0.492	41.892
Standard Deviation	(0.444)	(0.417)	(0.430)	(0.426)	(0.491)	(0.495)	(0.088)	(21.502)
Observations	72248	72248	72248	72248	72248	72248	72248	72248
Key Stage 3								
Mean	-0.192	0.205	0.008	-0.008	0.017	-0.018	0.483	186.759
Standard Deviation	(0.398)	(0.399)	(0.372)	(0.373)	(0.404)	(0.404)	(0.050)	(58.550)
Observations	10415	10415	10415	10415	10415	10415	10415	10415
Key Stage 4								
Mean	-0.191	0.198	-0.014	0.015	-0.014	0.015	0.488	182.263
Standard Deviation	(0.403)	(0.385)	(0.399)	(0.407)	(0.406)	(0.421)	(0.050)	(61.344)
Observations	6681	6681	6681	6681	6681	6681	6681	6681

Notes: Unit of comparison is the within school , within key stage cohort. The summary statistics for the mean scores at the key stage are generate using weighted values as described in the methodology, whilst those for the proportion of the cohort that is female and the size of cohort within the school are weighted using the inverse of the number of cohorts within schools observed at each key stage. Key stage 1 is not formally assessed for science.

Table 4 Summary statistics – small primary schools broken down by key stage.

	Mean score for males in English	**Mean scores for females in English**	**Mean scores for males in mathematics**	**Mean scores for females in mathematics**	**Mean scores for males in science**	**Mean scores for females in Science**	**Proportion of the cohort that is female**	**Size of cohort within school**
Key Stage 1 in small primary schools								
Mean	-0.129	0.198	0.035	0.023			0.490	22.085
Standard Deviation	(0.476)	(0.443)	(0.481)	(0.453)			(0.111)	(5.545)
Observations	16681	16681	16681	16681			16681	16681
Key Stage 2 in small primary schools								
Mean	-0.097	0.220	0.083	0.032	0.039	0.034	0.494	22.070
Standard Deviation	(0.503)	(0.473)	(0.490)	(0.481)	(0.543)	(0.548)	(0.111)	(5.334)
Observations	13408	13408	13408	13408	13408	13408	13408	13408

Notes: Unit of comparison is the within school , within key stage cohort. The summary statistics for the mean scores at the key stage are generate using weighted values as described in the methodology, whilst those for the proportion of the cohort that is female and the size of cohort within the school are weighted using the inverse of the number of cohorts within schools observed at each key stage. Key stage 1 is not formally assessed for science. A small primary school is defined as one that has 30 or fewer pupils in all observed years.

Table 5 Results for all schools and levels pooled

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
All levels and schools pooled						
Proportion of the within-school cohort that is female	-0.066***	0.011	0.020	0.030**	0.025	0.043**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.018)	(0.018)
Observations	137083	137083	137083	137083	76003	76003
Adjusted R-squared	0.02	0.02	0.04	0.05	0.08	0.11
Estimate of the effect of a 1 point increase in the peer ability due to the change in the gender make-up.	-0.227	0.037	-1.077	-1.611	-1.992	-3.385

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Year and exam dummies are also included. Standard errors are clustered at school level.

Table 6 Testing the linearity of the pooled regressions

	English		**Mathematics**		**Science**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Proportion of pupils that is female (1)	-0.061***	0.018	0.037*	0.037*	0.071**	0.080***
	(0.022)	(0.022)	(0.021)	(0.021)	(0.031)	(0.031)
(1) interacted with 2^{nd} quartile	0.054	0.052	-0.086	-0.088	0.080	-0.034
dummy (2)	(0.080)	(0.078)	(0.066)	(0.062)	(0.086)	(0.084)
(1) interacted with 3^{rd} quartile	0.004	-0.055	-0.022	-0.051	-0.196*	-0.134
dummy (3)	(0.096)	(0.095)	(0.081)	(0.081)	(0.108)	(0.108)
(1) interacted with 4^{th} quartile	-0.020	-0.015	-0.022	0.004	-0.085*	-0.054
dummy (4)	(0.037)	(0.036)	(0.033)	(0.033)	(0.050)	(0.049)
Observations	137083	137083	137083	137083	76003	76003
R-squared	0.02	0.01	0.03	0.03	0.07	0.09
P> F test statistic (2)=(3)=(4)=0	0.8220	0.8999	0.3610	0.3240	0.2831	0.4906

Notes: Robust standard errors in parentheses *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Year and exam dummies are also included. Standard errors are clustered at school level. Quartile dummies are based on 4 quantiles based on the proportion of pupils who are female.

Table 7 The effect of a more female peer group, broken down by primary and secondary schools and key stages

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Pooled primary schools						
Proportion of the within-school	-0.050***	0.008	0.040***	0.030***	0.061***	0.048***
cohort that is female	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	124297	124297	124297	124297	63217	63217
Adjusted R-squared	0.03	0.02	0.04	0.05	0.10	0.13
Pooled secondary schools						
Proportion of the within-school	-0.131***	0.018	-0.056	0.034	-0.049	0.028
cohort that is female	(0.050)	(0.049)	(0.042)	(0.043)	(0.044)	(0.043)
Observations	12786	12786	12786	12786	12786	12786
Adjusted R-squared	0.02	0.01	0.05	0.05	0.07	0.09
Key Stage 1						
Proportion of the within-school	-0.077***	-0.022	0.049***	0.028*	**N/A**	**N/A**
cohort that is female	(0.015)	(0.014)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Adjusted R-squared	0.01	0.02	0.05	0.05		
Key Stage 2						
Proportion of the within-school	-0.024	0.041***	0.025*	0.029**	0.061***	0.048***
cohort that is female	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Adjusted R-squared	0.05	0.04	0.06	0.08	0.10	0.13
Key Stage 3						
Proportion of the within-school	-0.137**	0.059	-0.056	0.033	-0.038	0.028
cohort that is female	(0.068)	(0.072)	(0.037)	(0.037)	(0.043)	(0.042)
Observations	8332	8332	8332	8332	8332	8332
Adjusted R-squared	0.03	0.01	0.10	0.13	0.20	0.22
Key Stage 4						
Proportion of the within-school	-0.129**	-0.017	-0.070	0.019	-0.079	0.012
cohort that is female	(0.063)	(0.059)	(0.060)	(0.060)	(0.063)	(0.061)
Observations	4454	4454	4454	4454	4454	4454
Adjusted R-squared	0.01	0.02	0.06	0.04	0.01	0.01

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Year and exam dummies are also included. Standard errors are clustered at school level.

Table 8 **Estimated effect of a 1 point average increase in the ability of the peer group from the change in the gender make-up**

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Pooled primary schools	**-0.177**	0.027	**-1.120**	**-0.859**	**-5.912**	**-4.663**
	124297	124297	124297	124297	63217	63217
Pooled Secondary Schools	**-0.420**	0.059	1.669	-1.031	1.346	-0.765
	12786	12786	12786	12786	12786	12786
Key Stage 1	**-0.261**	-0.076	**-18.030**	**-10.201**		
	61080	61080	61080	61080		
Key Stage 2	-0.087	**0.149**	**-0.352**	**-0.416**	**-5.912**	**-4.663**
	63217	63217	63217	63217	63217	63217
Key Stage 3	**-0.459**	0.199	0.960	-0.567	0.658	-0.480
	8332	8332	8332	8332	8332	8332
Key Stage 4	**-0.381**	-0.051	6.983	-1.926	4.160	-0.650
	4454	4454	4454	4454	4454	4454

Notes: The coefficients estimated here are the estimated effect of a 1 point change in the ability of the peer group related to a change in the gender make-up of the peer effect. Coefficient in bold indicates that the corresponding estimate of a more female peer group is significantly different from 0 at the 10% significance level. Method of estimation is as per Hoxby (2000). See appendix for details.

Table 9 Results in the subset of small and large primary schools

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Large Primary Schools						
Proportion of the within-school	-0.032*	0.006	0.054***	0.021	0.079***	0.065***
cohort that is female	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.025)
Observations	55211	55211	55211	55211	29099	29099
Adjusted R-squared	0.03	0.03	0.06	0.07	0.13	0.16
Small Primary Schools						
Proportion of the within-school	-0.105***	0.015	0.014	0.056**	-0.020	0.007
cohort that is female	(0.023)	(0.022)	(0.024)	(0.023)	(0.036)	(0.034)
Observations	26030	26030	26030	26030	11732	11732
Adjusted R-squared	0.01	0.01	0.02	0.03	0.05	0.07
Key Stage 1 in small primary schools						
Proportion of the within-school	-0.145***	-0.005	0.022	0.067**	**N/A**	**N/A**
cohort that is female	(0.029)	(0.027)	(0.032)	(0.030)		
Observations	14298	14298	14298	14298		
Adjusted R-squared	0.01	0.01	0.03	0.03		
Key Stage 2 in small primary schools						
Proportion of the within-school	-0.055	0.040	0.001	0.041	-0.020	0.007
cohort that is female	(0.036)	(0.035)	(0.035)	(0.033)	(0.036)	(0.034)
Observations	11732	11732	11732	11732	11732	11732
Adjusted R-squared	0.03	0.03	0.03	0.04	0.05	0.07

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. In square brackets are the translated effects of the coefficients of the exogenous change in peer tests scores that occurs from a change in the gender make-up of the peer group. Method is weighted least squares. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 10 Comparisons of full sample and credibly random sample

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
All levels and schools pooled						
Schools that have apparent random changes in gender make-up	-0.068***	0.014	-0.002	0.035**	0.027	0.063***
	(0.018)	(0.017)	(0.017)	(0.016)	(0.025)	(0.024)
Observations	72684	72684	72684	72684	40176	40176
All Schools	-0.066***	0.011	0.020	0.030**	0.025	0.043**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.018)	(0.018)
Observations	137083	137083	137083	137083	76003	76003
Primary Schools						
Schools that have apparent random changes in gender make-up	-0.053***	0.005	0.029*	0.037**	0.070***	0.066***
	(0.015)	(0.014)	(0.015)	(0.014)	(0.021)	(0.021)
Observations	65870	65870	65870	65870	33362	33362
All Schools	-0.050***	0.008	0.040***	0.030***	0.061***	0.048***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	124297	124297	124297	124297	63217	63217
Secondary Schools						
Schools that have apparent random changes in gender make-up	-0.124*	0.041	-0.111**	0.034	-0.055	0.051
	(0.065)	(0.063)	(0.055)	(0.055)	(0.058)	(0.055)
Observations	6814	6814	6814	6814	6814	6814
All Schools	-0.131***	0.018	-0.056	0.034	-0.049	0.028
	(0.050)	(0.049)	(0.042)	(0.043)	(0.044)	(0.043)
Observations	12786	12786	12786	12786	12786	12786

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 11 Comparisons of full sample and credibly random sample, broken down by key stage

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Key Stage 1						
Schools that have apparent random changes in gender make-up	-0.081***	-0.019	0.032	0.043**	**N/A**	**N/A**
	(0.020)	(0.018)	(0.021)	(0.020)		
Observations	32508	32508	32508	32508		
All Schools	-0.077***	-0.022	0.049***	0.028*	**N/A**	**N/A**
	(0.015)	(0.014)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Key Stage 2						
Schools that have apparent random changes in gender make-up	-0.023	0.031	0.024	0.030	0.070***	0.066***
	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
Observations	33362	33362	33362	33362	33362	33362
All Schools	-0.024	0.041***	0.025*	0.029**	0.061***	0.048***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Key Stage 3						
Schools that have apparent random changes in gender make-up	-0.147*	0.091	-0.105**	0.008	-0.084	0.015
	(0.089)	(0.093)	(0.049)	(0.048)	(0.057)	(0.056)
Observations	4456	4456	4456	4456	4456	4456
All Schools	-0.137**	0.059	-0.056	0.033	-0.038	0.028
	(0.068)	(0.072)	(0.037)	(0.037)	(0.043)	(0.042)
Observations	8332	8332	8332	8332	8332	8332
Key Stage 4						
Schools that have apparent random changes in gender make-up	-0.111	0.002	-0.139*	0.045	-0.057	0.070
	(0.083)	(0.078)	(0.081)	(0.080)	(0.082)	(0.080)
Observations	2358	2358	2358	2358	2358	2358
All Schools	-0.129**	-0.017	-0.070	0.019	-0.079	0.012
	(0.063)	(0.059)	(0.060)	(0.060)	(0.063)	(0.061)
Observations	4454	4454	4454	4454	4454	4454

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Key stage 1 is not formally assessed for science Standard errors are clustered at school level

Table 12 Comparisons of full sample and credibly random sample in small and large primary schools

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Large Primary Schools						
Schools that have apparent random changes in gender make-up	-0.020	0.034	0.053**	0.045**	0.092***	0.098***
	(0.022)	(0.021)	(0.023)	(0.022)	(0.032)	(0.033)
	29627	29627	29627	29627	15491	15491
All Schools	-0.032*	0.006	0.054***	0.021	0.079***	0.065***
	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.025)
	55211	55211	55211	55211	29099	29099
Small Primary Schools						
Schools that have apparent random changes in gender make-up	-0.117***	-0.024	-0.005	0.033	-0.036	-0.010
	(0.031)	(0.030)	(0.032)	(0.029)	(0.047)	(0.045)
	13813	13813	13813	13813	6181	6181
All Schools	-0.105***	0.015	0.014	0.056**	-0.020	0.007
	(0.023)	(0.022)	(0.024)	(0.023)	(0.036)	(0.034)
	26030	26030	26030	26030	11732	11732
Key Stage 1 in small primary schools						
Schools that have apparent random changes in gender make-up	-0.153***	-0.018	-0.007	0.061	N/A	N/A
	(0.039)	(0.036)	(0.042)	(0.040)		
	7632	7632	7632	7632		
All Schools	-0.145***	-0.005	0.022	0.067**	N/A	N/A
	(0.029)	(0.027)	(0.032)	(0.030)		
	14298	14298	14298	14298		
Key Stage 2 in small primary schools						
Schools that have apparent random changes in gender make-up	-0.070	-0.034	-0.002	-0.006	-0.036	-0.010
	(0.050)	(0.049)	(0.047)	(0.043)	(0.047)	(0.045)
	6181	6181	6181	6181	6181	6181
All Schools	-0.055	0.040	0.001	0.041	-0.020	0.007
	(0.036)	(0.035)	(0.035)	(0.033)	(0.036)	(0.034)
	11732	11732	11732	11732	11732	11732

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 13 OLS estimation including socioeconomic factors within primary schools

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Key Stage 1						
Proportion of the within-school cohort that is female	-0.073***	-0.025*	0.052***	0.026*	N/A	N/A
	(0.014)	(0.013)	(0.016)	(0.015)		
Proportion of males that receive FSM within cohort	-0.465***	0.023*	-0.389***	0.012		
	(0.015)	(0.014)	(0.017)	(0.016)		
Proportion of females that receive FSM within cohort	0.021	-0.415***	0.013	-0.357***		
	(0.014)	(0.013)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Adjusted R-squared	0.03	0.04	0.06	0.06		
Key Stage 2						
Proportion of the within-school cohort that is female	-0.019	0.035**	0.029*	0.024	0.065***	0.042***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Proportion of males that receive FSM within cohort	-0.434***	-0.015	-0.374***	0.009	-0.370***	0.009
	(0.016)	(0.015)	(0.016)	(0.015)	(0.017)	(0.016)
Proportion of females that receive FSM within cohort	0.026*	-0.399***	0.007	-0.376***	0.014	-0.386***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Adjusted R-squared	0.07	0.05	0.07	0.09	0.11	0.14

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. In square brackets are the translated effects of the coefficients of the exogenous change in peer tests scores that occurs from a change in the gender make-up of the peer group. Method is weighted least squares. FSM is free school meals. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 14 Value added

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
All levels and schools pooled						
Proportion of the within-school cohort	-0.016	0.014	0.016	0.067***	-0.042	0.066*
that is female	(0.028)	(0.028)	(0.022)	(0.023)	(0.038)	(0.037)
Observations	25908	25908	25908	25908	12306	12306
Adjusted R-squared	0.06	0.03	0.13	0.20	0.33	0.43
Pooled Secondary Schools						
Proportion of the within-school cohort	-0.101**	-0.033	0.005	0.086***	-0.042	0.066*
that is female	(0.043)	(0.044)	(0.027)	(0.031)	(0.038)	(0.037)
Observations	12306	12306	12306	12306	12306	12306
Adjusted R-squared	0.06	0.03	0.22	0.33	0.33	0.43
Key Stage 2						
Proportion of the within-school cohort	0.089***	0.071**	0.031	0.044		
that is female	(0.031)	(0.030)	(0.033)	(0.033)		
Observations	13602	13602	13602	13602		
Adjusted R-squared	0.03	0.01	0.06	0.06		
Key Stage 3						
Proportion of the within-school cohort	-0.084	0.054	-0.011	0.108***	-0.075	0.086*
that is female	(0.070)	(0.073)	(0.036)	(0.037)	(0.051)	(0.049)
Observations	8284	8284	8284	8284	8284	8284
Adjusted R-squared	0.08	0.04	0.27	0.41	0.38	0.46
Key Stage 4						
Proportion of the within-school cohort	-0.119	-0.194*	0.033	0.016	0.002	0.010
that is female	(0.098)	(0.101)	(0.048)	(0.045)	(0.058)	(0.060)
Observations	4022	4022	4022	4022	4022	4022
Adjusted R-squared	0.02	0.03	0.08	0.11	0.22	0.34

Notes: Dependent variable is the change in mean value added score from one key stage to the next within school cohort for male (female) pupils. Only pupils who remain in the same school from key stage 1 to key stage 2 and key stage 3 to key stage 4 are included, whilst pupils who change schools between key stage 3 and 4 are included. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Each cell represents a separate regression. Science has no regressions at key stage 2 as the pupils are not formally assessed at key stage 1 for science. Standard errors are clustered at school level

Table 15 Value added in small primary schools

	English		Mathematics	
	Male	**Female**	**Male**	**Female**
Small school Key Stage 2				
Proportion of the within-school cohort that is female	0.072	0.017	-0.055	0.011
	(0.064)	(0.059)	(0.067)	(0.058)
Observations	2848	2848	2848	2848
Adjusted R-squared	0.03	0.01	0.05	0.04
Large school Key Stage 2				
Proportion of the within-school cohort that is female	0.103	0.085	0.103	0.085
	(0.064)	(0.054)	(0.064)	(0.054)
Observations	4938	4938	4938	4938
Adjusted R-squared	0.03	0.01	0.03	0.01

Notes: Dependent variable is the change in mean value added score from one key stage to the next within school cohort for male (female) pupils. Only pupils who remain in the same school from key stage 1 to key stage 2 and key stage 3 to key stage 4 are included, whilst pupils who change schools between key stage 3 and 4 are included. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Science has no regressions at key stage 2 as the pupils are not formally assessed at key stage 1 for science. Standard errors are clustered at school level