THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

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.

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

Steven Proud

January 2008

Working Paper No. 08/186

University of BRISTOL

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The Leverhulme Trust

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

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 follow several different mechanisms and are often difficult to estimate, due to unobserved selection. This paper builds on the work of Hoxby (2000) and uses exogenous changes in the proportion of girls within UK school cohorts to estimate the effect of a more female peer group. I include estimates of effects at a classroom level for schools that appear to contain only one class per cohort to estimate the direct effect of a peer group. Further, I examine if there is a differential effect of boys and girls with differing socioeconomic status, and also examine the effect of a more female peer group on a child's value added score. I find large significant negative effects of a more female peer group on boy's outcomes in English, whilst in maths and science, both boys and girls benefit from a more able peer group up until age 11.

Keywords: peer groups, education

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/

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. This paper utilizes the same strategy, 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].

Whilst the majority of the literature addresses the effects of either the ability of a peer group or the racial make-up, there is a smaller literature addressing the effects of the gender make-up of the peer group, with most studies addressing the effects of single sex schools against mixed sex schools. For example Marsh and Rowe (1996) find little effect of single sex classes, with male pupils feeling less favourable to single sex classes. In the UK, Malacova (2007) employs multilevel methodology, and finds an advantage for girls educated in single sex classrooms, but with this advantage

---

[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.

decreasing according to prior ability, whilst the advantage of single sex education decreased for boys according to the level of school selectiveness.[4]

Hoxby (2000), however, considers the effect of a change in the proportion of the peer group who are female, and uses multiple strategies to estimate the effect of gender and race make-up of the peer group on individual outcomes. The initial strategy makes use of her argument that there is a credibly exogenous variation in the distribution of female pupils across cohorts within a grade, and simply uses the raw proportion of girls as a measure of the peer group. By utilising this strategy she finds that if all of the peer effects operate through peer ability, then a 1 point increase in peer ability should lead to a rise in pupils scores by between 0.3 and 0.5 points in reading, and in maths a raise of between 1.7 and 6.8 points. However, she goes on to argue that this is far too large an effect to be credible, and so other mechanisms must be in operation. For example, Lavy and Schlosser (2007) use an individual level dataset of children in Israel to estimate the effects and mechanisms of a more female peer groups. They use a rich dataset of children's behaviour and their peers' perceptions of behaviour. They find significant positive effects on cognitive outcomes of a more female peer group. Furthermore, they find that the higher proportion of female peers lowers the amount of classroom disruption and violence, although they go on to argue that the greater cognitive outcomes are due to compositional effects rather than behavioural effects or alternatively ability spillover.

Meanwhile, Whitmore (2005) finds positive effects of a high fraction of girls in kindergarten through to the second grade on the outcomes of1 both boys and girls, whilst in the third grade, finds 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.

One of the mechanisms that may be affecting the outcomes of pupils by a change in the proportion of peers that is female is a change in the average ability of the peer group, as measured by prior performance in key stage examinations. In order to estimate this, there needs to be a difference in the outcomes of girls and boys.

---

[4] A more complete analysis of the advantages and disadvantages of single sex education is provided in Campbell and Sanders (2002)

Burgess et al (2004) consider possible reasons for the gender achievement gap at age 16 in English schools. They find that the gap is largely seen with girls outperforming boys in English, with very little difference in performance in maths and science between equivalent male and female students. Gorard et al (2001) propose that much of the gender difference between girls and boys has been misunderstood in previous research. They contest that in all subjects other than maths and science (for which there are few significant gender specific differences) that all pupils start at the beginning of education on an equal footing, and over time the apparent gender gap develops. This is partially driven by proportionately more girls gaining high grades and more boys achieving middle level grades than may be expected.

Machin and MacNally (2005) consider how the gender gap has evolved over time at the end of primary school and secondary education[5]. They find the largest gender gap in primary schools is with relation to girls outperforming boys in reading, although this has decreased since the introduction of the literacy hour. Furthermore, they find that the increase of coursework at GCSE has benefited girls, and increased the gender gap. This finding is corroborated by Younger and Warrington (1996), and they find that teachers are apparently more lenient with girls than boys, indicating differential treatment. However, Myhill (2002) contradicts this and shows that the gender gap increased when the proportion of coursework was decreased. Stobart et al (1992) also find evidence that coursework does not favour girls.

Within mathematics, Shibley, Hyde et al (1990) find male dominance, which is decreasing over time. This finding is corroborated by Kraemer (2000) who comments on the gender gap with males possessing superior skills in mathematics and non-verbal tasks, with even 2-year-old boys better able to build bridges with toy bricks than similar girls. Girls, however, have better language skills and are more aware of their feelings. Furthermore, 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 sex of the pupil. Male high achievers are far more likely to

---

[5] Primary school covers age 4 to 11, Secondary from 11 to 16. A further description is given in section 3

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).

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 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 discussing the methodology used in this paper. Section 3 examines the PLASC and NPD dataset used here, section 4 examines the summary statistics and section 5 discussed the results. Finally, section 6 offers discussion and conclusions.

2. 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

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

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. I later try to control for changes in the demographics by including a measure of relative deprivation in the school; that is the proportion of pupils who receive free school meals (FSM)[6]. There may be a possibility that female pupils with a low socioeconomic status have a different effect to females with a high socioeconomic status, so to try and control for this effect, I enter the proportion of male pupils receiving FSM and the proportion of female pupils receiving FSM separately.

The exogenous and constant variables, *X,* consists 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.

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

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.

---

[6] Free school meals are only available to "Children whose parents receive Income Support (IS); Income-based Job Seekers Allowance (IBJSA); support under Part VI of the Immigration and Asylum Act 1999; or Child Tax Credit, but who are not entitled to Working Tax Credit and whose annual income (as assessed by the Inland Revenue) that from 6 April 2005 does not exceed £13,910; or the Guaranteed Element of State Pension Credit are entitled to free school meals. Children who receive IS or IBJSA in their own right are also entitled to free school meals". Current eligibility criteria for free school meals from
http://www.parentscentre.gov.uk/educationandlearning/schoollife/schooladministration/schoolmealsandmilk/

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.

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

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, I take first differences across cohorts within a given key stage,

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

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

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, a pupil who attends a school with a large proportion of pupils who are female may not experience this grouping within the 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 will later show 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.

Thus far, I have simply considered using the levels that students receive from examinations at ages 7, 11, 14 and 16. These levels are highly correlated with other, unobserved, factors such as family background and neighbourhood affluence. In order to try to control for this, I also examine a value added score within subject. The value added measure I use is simply the test score achieved by an individual pupil at one key stage subtracted from the score obtained at the subsequent key stage. For instance, the value added at age 11 is simply the test score at age 7 subtracted from the test score at age 11[7]. In order to examine the effects of a more female peer group, I would like pupils to remain in the treatment group for the whole period between examinations. Due to the structure of schools in England almost all pupils (98.6%) have changed schools between Key Stage 2 and 3, whilst few pupils change from Key Stage 3 to 4 (3.1%). Wilson (2003) shows that there is a low correlation between test scores and value added, and thus the effect of school level inputs may be better viewed using this value added score.

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.

---

[7] There are other methods of calculating value added.

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.

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.

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

(6)

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

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:

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

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

3. Data

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.

Pupils are assessed at 4 key stages through their school careers, at ages 7, 11, 14 and 16. 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.

Pupils are examined in reading, writing and mathematics at KS1, English maths 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.

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 simply consider an A* to be worth 8 points and a G to be worth 1 point.

At present, it is not possible to observe any pupils through the entire assessment process in schools with the data in PLASC, 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.

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.

Infant schools cover the first three years of primary, from age 4 to age 7. In infant schools since the start of the 2001/2002 academic year there is a legal requirement that there should not be more than 30 pupils to a qualified teacher (the Education (Infant Class Sizes) Regulations 1998). Effectively, this means that the maximum class size in infant's school is 30. However, according to Smithers (2006), Department for Education and Skills statistics show that 29,000 pupils, or 2.1% of all infant age pupils, are taught in classes of 31 or more. However, this is qualified since

some of the large class-sizes are approved due to children being moved into a school's area after the start of the school-year or when the local authority has placed a child with special needs into a school. Figure 2 shows the distribution 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.

Sample Selection

In order to control for endogeneities caused by selection, I only consider here schools in non-selective local authorities[8] (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 select pupils based on ability, the non-selective schools are left with a non-random selection of pupils. Furthermore, I only include community, community special, voluntary aided, voluntary controlled,

---

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

foundation, foundation special 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 1st to the 99th 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 1st to 99th 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. Summary Statistics

Table 3 shows summary statistics for all of the data, in various forms. Firstly, as a general overview, summary statistics of all of the data pooled is shown, then by

primary and secondary schools, then by key stage, and finally, examining the differences between small and large primary schools. The scores from English, maths 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 from the raw 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 maths are.

Since all of the key stage results are based on means of normalised results centred at 0 with standard deviation, 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 maths and science, there is little or no difference, although boys do have a very slight advantage in maths.

At key stage 1, in English, girls have a significant advantage in English, whilst there is little difference between the genders in maths. At key stage 2, in English, there is still a significant advantage for girls in English, whilst in maths 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 maths 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 maths 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 (indicating a nature of the English school system, with secondary schools generally much larger institutions than primary schools). This may make inferences at a school level much harder at the secondary level much more difficult 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 propagate down to the classroom level, whether due to ability setting or some other mechanism.

Furthermore, it appears that in general, boys and girls on average perform better in a small (one classroom) primary school than in a larger one.

5. Results

In looking at the results, I start by looking at a specification that includes all of the schools, given sample selection, and all of the available key stages, followed by tests of linearity of the specification. This is then followed by specifications just 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 the results with results from a subset of the sample that only contains schools that appear to have completely random changes in the gender make-up from year to year. I then consider the effects of a measure of poverty for boys and girls. 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.

Results in all schools

Table 4 shows regression results for all schools in English, maths and science. The initial specification includes all schools and levels, and I estimate equation (5) using the weightings described in equation (7):

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

where p is the proportion of pupils in the cohort within the school that are 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 maths and science; both girls and 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 English scores for boys by approximately 0.015 standard deviations, a rise in

maths of approximately 0.006 standard deviations and a rise of 0.007 standard deviations for science. For girls, this would lead to an increase in maths score by 0.007 standard deviations and in science by 0.010 standard deviations.

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 are female. Figure 8 shows adjusted variable plots for the pooled regressions in English, maths and science. In the graphs, the x axis represents the 50 quantiles of the proportion of pupils that are 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. These all appear to follow a fairly linear pattern, other than for females in English, which does not seem to follow any real pattern. However, it is necessary to check this.

Table 5 reports the results of regressions for equation (6) using the pooled specification above, including terms interacting the proportion of pupils that are female with the quartile that it is in. Using an F-test, in all of the subjects for all of the pupils, the test does not reject the null that the coefficients on all of these interaction terms are equal, and equal to zero, so does not reject the null of linearity.

Table 4 also 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 are 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 maths and science, both male and female pupils see a significant and positive effect of a more female peer group in primary schools. If these results are 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, I find large negative effects of an exogenous increase in the peer group ability, which are of a much larger magnitude than could be credibly expected. The same effects are seen for maths within key stages 1 and 2. As these effects are of an unbelievable magnitude, I do not consider these further.

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.

Small and large primary schools

Table 6 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 than 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.

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, tables 7 and 8 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 7, for English, there are no major differences between the full sample and the apparently random sample. In mathematics, there is a major difference between the results in all levels and schools pooled, and for secondary schools, and within key stages 3 and 4 for males. However, whilst these differences are large, they 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, the difference is not significant.

Table 8 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 result is significantly different from 0, so it does not affect my results.

Socioeconomic status

Table 9 shows a breakdown by key stage of results using a measure of socioeconomic status of the school, the proportion of male and female pupils who receive free school meals within the cohort. Introducing this measure has no significant effect on the coefficient on the proportion of pupils in the cohort that are female in any subject, in any assessment level. Furthermore, the gender specific socioeconomic status has a significant negative effect on outcomes. For instance, boys in a cohort with a large proportion of males with free school meals do significantly worse in all subjects, and similarly for girls. However, in English at key stage 1, there is a small anomaly. The proportion of male FSM pupils in the cohort in the school actually has a small significant positive impact on females' results.

For the gender specific socioeconomic status, the effect seen is constant through primary school, and then increases through secondary school, with the effects for male and female pupils not significantly different.

Since there is no significant change in the coefficient on the proportion of pupils in the school-cohort that are female, I conclude that the socioeconomic status of the school has the same effect on boys and girls and the omission of this variable is not creating any bias in the results.

Value Added Results.

Table 10 shows the results of the estimation of equation (5), 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 maths 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 10 with those from table 4, 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 are female.

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 that are obtained show large 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. Hoxby (2000) uses the effect of a change in the proportion of pupils that are female to try and estimate the effect of a credible change in the ability of the peer group, although she does qualify the results with the proviso that her results are of a too high a magnitude to be plausible. However, considering the results I obtain here, due to the considerably higher scores in English achieved by female students, and the slightly higher scores in maths and science achieved by male students in primary schools (see table 3), I find large negative effects of a more able peer group for boys in English at all stages of education, and in maths and science for

both boys and girls in primary schools, in contradiction with current established literature. For example Lefgren (2004) finds significant positive effects of a more able peer group, a finding that is backed up by Zimmerman (2003).

This is not the whole story. In mathematics and science, the results of male and female pupils is very closely matched, and so a very large change in the gender make-up of the peer group is required in order to have any noticeable change in the ability of the peer group, meaning that the estimates of the effect of a change in the ability of the peer group would require such a large change in the gender make-up of the cohort that we cannot take the implied result seriously.

Lavy and Schlosser (2007) offer a different explanation. They show that if the results were driven by increases in ability of the peer group, then in contradiction with the established literature, this would lead to a decrease in the ability of the peers. 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 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, but for many this is just an image. Furthermore, there is also evidence that boys require more behavioural management than girls. According to the official data in PLASC, 70.9 percent of children with statements are boys, and further 65.4% of all children with 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 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[9], the negative effects for boys in English disappear, but apparent positive effects are seen in maths and science which are not seen in the small schools' case.

Overall, the results imply that in primary schools at least, boys would benefit greatly from being taught English in single sex classes, which would have little effect on girls' outcomes, whilst in maths and science, different policies would benefit boys and girls: boys would be better off in a more female classroom, whilst girls would be better off in an all female classroom. This last conclusion, however, is in contradiction to Asthana (2006), who quotes the findings of Alan Smithers' research, claiming that there is no advantage of teaching girls in a single sex environment, contradicting the long-held view that in schools girls are distracted by boys in the classroom, and other arguments that girls and boys brains develop differently and thus require different emphases in teaching. Smithers' research examines data from across the world, and finds little impact of consistent superior performance in single sex schools.

Smithers' research is almost in direct contradiction to Younger et al (2005) who when examining whole school approaches to raising boys' achievement consider the effects of single-sex classes. They find evidence that "girls and boys feel more at ease in such classes, feel more able to interact with learning and to show real interest without inhibition and often achieve more highly as a result"[10] Thus, whilst my results back up Younger et al (2005) for English and for girls in maths and science, single sex classes in maths and science for boys would have a detrimental effect. Furthermore, Jackson (2002) finds that single sex classes are likely to have positive effects for girls, but male only classes may exacerbate the macho male cultures inherent in schools.

Whilst it is not possible to generalise to the limit of single sex classrooms, the results obtained imply 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

---

[9] Schools with more than 30 pupils in the cohort.
[10] Younger et al (2005) page 12

proportion of girls for both boys and girls, implying that a mix of the genders is optimal in both maths and science.

References

Asthana, A. (2006). Single-sex schools 'no benefit for girls'. The Observer. London**:** 1.

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): 209-29.

Campbell, P. B. and J. Sanders (2002). "Challenging the system: Assumptions and data behind the push for single-sex schooling" in Doing Gender in Policy and Practice: Perspectives on Single Sex and Coeducational Schooling. A. Datnow and L. Hubbard. New York, Routledge Falmer**:** 31-46.

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

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." London, Routledge Falmer.

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 A.B. Sorensen, (1987) *Ability Grouping and Sex Differences in Mathematics Achievement.* Sociology of Education,. **60**(April): 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.", NBER working paper 12251

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation., NBER working paper 7867

Hoxby, C. (2002). "The Power of Peers." Education Next **2.**

Jackson, C. (2002). "Can single-sex classes in co-educational schools enhance the learning experiences of girls and/or boys? An exploration of pupils' perceptions." British Educational Research Journal **28**(1): 37-48.

Kraemer, S. (2000), "The Fragile Male*"*. British Medical Journal,. **321**: 1609-1612.

Lavy, V. and A. Schlosser (2007). "Mechanisms and Impacts of Gender Peer Effects at School", NBER Working Papers: 13292

Lefgren, L. (2004). "Educational Peer Effects and the Chicago Public Schools." Journal of Urban Economics **56**(2): 169-91.

Machin, S. and S. McNally (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. London.

Malacova, E. (2007). "Effect of single-sex education on progress in GCSE." Oxford Review of Education **33**(2): 233-259.

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

Marsh, H. W. and Rowe, K. J. (1996). "The effects of single-sex and mixed-sex mathematics classes within a coeducational school: A reanalysis and comment." Australian Journal of Education **40**(2): 147-161.

Myhill, D. (2002) "Bad boys and good girls?: patterns of interaction and responses in whole-class teaching", British Educational Research Journal, **28**(3): 339– 352.

Shibley-Hyde, J., Fennema, E., and Lamon, S. J., (1990) "Gender Differences in Mathematics Performance: A Meta-Analysis", Psychological Bulletin Copyright 1990 by the American Psychological Association, Inc. **107**(2): 139-155

Smithers, R. (2006). "Primary schools have 29,000 pupils in classes of over 30". The Guardian. London.

Stobart, G., J. Elwood, and M. Quinlan, (1992) "Gender Bias in Examinations: how equal are the opportunities". British Educational Research Journal, **18**(3): 261-276

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.

Wilson, D. (2004). "Which ranking? The impact of a 'value-Added' measure of secondary school performance." Public Money & Management **24**(1): 37-45.

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

Younger, M. and M. Warrington (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., Williams, J., (1999) "The Gender Gap and Classroom Interactions: reality and rhetoric?" British Journal of Sociology of Education, **20**(3): 325-341

Zimmerman, D.-J. (2003). ": Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.": Review-of-Economics-and-Statistics. **85**(1): 9-23

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	962,039 (100%)	1,429,998 (100%)	1,115,952 (100%)
Pupils at same school	711,014 (73.9%)	19,748 (1.4%)	1,081,511 (96.9%)
Pupils at different school	251,025 (26.1%)	1,410,250 (98.6%)	34,411 (3.1%)

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 are female	Size of cohort within school
Pooled Specification								
Mean	-0.17	0.18	0.01	-0.01	0.00	0.00	0.49	111.54
Standard Deviation	0.42	0.40	0.41	0.41	0.44	0.44	0.07	84.59
Observations	168815	168815	168815	168815	94132	94132	168815	168815
Primary Schools								
Mean	-0.15	0.16	0.02	-0.02	0.01	-0.01	0.49	40.68
Standard Deviation	0.44	0.41	0.43	0.42	0.49	0.50	0.09	20.38
Observations	150675	150675	150675	150675	75992	75992	150675	150675
Secondary Schools								
Mean	-0.19	0.20	0.00	0.00	0.00	0.00	0.48	182.39
Standard Deviation	0.41	0.40	0.40	0.40	0.41	0.42	0.05	62.09
Observations	18140	18140	18140	18140	18140	18140	18140	18140
Key Stage 1								
Mean	-0.16	0.16	0.00	0.00	N/A	N/A	0.49	39.44
Standard Deviation	0.43	0.4	0.43	0.41	N/A	N/A	0.09	19.10
Observations	74683	74683	74683	74683	N/A	N/A	74683	74683
Key Stage 2								
Mean	-0.15	0.15	0.03	-0.03	0.01	-0.01	0.49	41.92
Standard Deviation	0.45	0.45	0.42	0.43	0.43	0.49	0.09	21.50
Observations	75992	75992	75992	75992	75992	75992	75992	75992
Key Stage 3								
Mean	-0.19	0.20	0.01	-0.01	0.02	-0.02	0.48	183.09
Standard Deviation	0.41	0.41	0.39	0.39	0.42	0.42	0.05	62.59
Observations	11135	11135	11135	11135	11135	11135	11135	11135
Key Stage 4								
Mean	-0.19	0.20	-0.01	0.01	-0.01	0.01	0.49	181.69
Standard Deviation	0.41	0.39	0.4	0.41	0.41	0.42	0.05	61.58
Observations	7005	7005	7005	7005	7005	7005	7005	7005
Small Primary Schools								
Mean	-0.12	0.2	0.05	0.02	0.03	0.03	0.49	20.86
Standard Deviation	0.50	0.46	0.50	0.47	0.56	0.56	0.11	5.60
Observations	31601	31601	31601	31601	14080	14080	31601	31601
Large Primary Schools								
Mean	-0.15	0.16	0.02	-0.02	0.01	-0.01	0.49	57.58
Standard Deviation	0.42	0.39	0.41	0.39	0.47	0.47	0.07	18.39
Observations	67025	67025	67025	67025	35000	35000	67025	67025
Key Stage 1 in small primary schools								
Mean	-0.13	0.20	0.03	0.02	N/A	N/A	0.49	20.83
Standard Deviation	0.48	0.44	0.49	0.45	N/A	N/A	0.11	5.69
Observations	17521	17521	17521	17521	17521	17521	17521	17521
Key Stage 2 in small primary schools								
Mean	-0.11	0.21	0.07	0.03	0.03	0.03	0.49	20.91
Standard Deviation	0.52	0.49	0.51	0.49	0.56	0.56	0.11	5.49
Observations	14080	14080	14080	14080	14080	14080	14080	14080

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 are female and the size of cohort within the school are weighted using the number of cohorts within schools observed at each key stage. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level

Table 4 Results in all schools.

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
All levels and schools pooled						
Proportion of the within-school cohort that are female	**-0.064***	**0.010**	**0.023***	**0.030****	**0.029***	**0.042****
	(0.013)	(0.012)	(0.012)	(0.012)	(0.018)	(0.017)
	[-0.201***]	[0.031]	[-1.424*]	[-1.887**]	[-2.822*]	[-4.024**]
Observations	144085	144085	144085	144085	80071	80071
Adjusted R-squared	0.02	0.02	0.04	0.05	0.08	0.11
Pooled primary schools						
Proportion of the within-school cohort that are female	**-0.051***	**0.006**	**0.040***	**0.031***	**0.056***	**0.044***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.015)	(0.015)
	[-0.18***]	[0.02]	[-1.11***]	[-0.86***]	[-5.17***]	[-4.05***]
Observations	130507	130507	130507	130507	66493	66493
Adjusted R-squared	0.03	0.02	0.04	0.05	0.10	0.13
Pooled Secondary Schools						
Proportion of the within-school cohort that are female	**-0.118****	**0.019**	**-0.040**	**0.035**	**-0.028**	**0.032**
	(0.048)	(0.048)	(0.040)	(0.041)	(0.043)	(0.042)
	[-0.37**]	[0.06]	[1.54]	[-1.35]	[0.91]	[-1.22]
Observations	13578	13578	13578	13578	13578	13578
Adjusted R-squared	0.02	0.01	0.05	0.05	0.07	0.09
Key Stage 1						
Proportion of the within-school cohort that are female	**-0.076***	**-0.020**	**0.049***	**0.030****	**N/A**	**N/A**
	(0.014)	(0.013)	(0.016)	(0.015)		
	[-0.26***]	[-0.07]	[-12.89***]	[-7.84**]		
Observations	64014	64014	64014	64014		
Adjusted R-squared	0.01	0.02	0.05	0.05		
Key Stage 2						
Proportion of the within-school cohort that are female	**-0.025***	**0.035****	**0.025***	**0.028***	**0.056***	**0.044***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
	[-0.09*]	[0.13**]	[-0.35*]	[-0.40*]	[-5.17***]	[-4.05***]
Observations	66493	66493	66493	66493	66493	66493
Adjusted R-squared	0.05	0.04	0.06	0.08	0.10	0.13
Key Stage 3						
Proportion of the within-school cohort that are female	**-0.127***	**0.049**	**-0.035**	**0.033**	**-0.015**	**0.035**
	(0.065)	(0.069)	(0.036)	(0.036)	(0.041)	(0.040)
	[-0.41*]	[0.16]	[0.76]	[-0.70]	[0.31]	[-0.71]
Observations	8908	8908	8908	8908	8908	8908
Adjusted R-squared	0.03	0.01	0.10	0.13	0.20	0.23
Key Stage 4						
Proportion of the within-school cohort that are female	**-0.112***	**-0.007**	**-0.056**	**0.023**	**-0.055**	**0.017**
	(0.062)	(0.058)	(0.059)	(0.058)	(0.061)	(0.060)
	[-0.33*]	[-0.02]	[9.40]	[-3.96]	[5.59]	[-1.69]
Observations	4670	4670	4670	4670	4670	4670
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. 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. 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 5 Testing the linearity of the pooled regressions

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Proportion of pupils that are female (1)	**-0.067***	**0.016**	**0.034***	**0.042****	**0.062****	**0.081***
	(0.022)	(0.021)	(0.020)	(0.020)	(0.030)	(0.030)
(1) interacted with 2nd quartile dummy (2)	**0.043**	**0.042**	**-0.058**	**-0.075**	**0.100**	**-0.019**
	(0.078)	(0.076)	(0.064)	(0.060)	(0.084)	(0.082)
(1) interacted with 3rd quartile dummy (3)	**0.042**	**-0.034**	**-0.011**	**-0.067**	**-0.153**	**-0.133**
	(0.094)	(0.093)	(0.079)	(0.079)	(0.105)	(0.106)
(1) interacted with 4th quartile dummy (4)	**-0.006**	**-0.015**	**-0.015**	**-0.007**	**-0.068**	**-0.064**
	(0.036)	(0.035)	(0.032)	(0.032)	(0.048)	(0.048)
Observations	144085	144085	144085	144085	80071	80071
R-squared	0.02	0.01	0.03	0.03	0.07	0.09
P> F test statistic (2)=(3)=(4)=0	**0.8031**	**0.9504**	**0.6657**	**0.3461**	**0.4031**	**0.4732**

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.

Table 6 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 cohort that are female	-0.033**	0.006	0.052***	0.023	0.077***	0.061**
	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.024)
	[-0.11**]	[0.02]	[-1.75***]	[-0.76]	[-7.29***]	[-5.76**]
Observations	58075	58075	58075	58075	30625	30625
Adjusted R-squared	0.03	0.03	0.06	0.07	0.13	0.16
Small Primary Schools						
Proportion of the within-school cohort that are female	-0.108***	0.015	0.015	0.059***	-0.032	0.010
	(0.022)	(0.021)	(0.024)	(0.022)	(0.035)	(0.033)
	[-0.41***]	[0.06]	[-0.30]	[-1.18***]	[1.41]	[-0.43]
Observations	27338	27338	27338	27338	12320	12320
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 cohort that are female	-0.146***	-0.002	0.027	0.073**	N/A	N/A
	(0.028)	(0.026)	(0.031)	(0.029)		
	[-0.56***]	[-0.01]	[-0.77]	[-2.05**]		
Observations	15018	15018	15018	15018		
Adjusted R-squared	0.01	0.01	0.03	0.03		
Key Stage 2 in small primary schools						
Proportion of the within-school cohort that are female	-0.059*	0.037	-0.004	0.041	-0.032	0.010
	(0.035)	(0.034)	(0.034)	(0.032)	(0.035)	(0.033)
	[-0.22*]	[0.14]	[0.06]	[-0.57]	[1.41]	[-0.43]
Observations	12320	12320	12320	12320	12320	12320
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 7 Comparison – Ordered and Random

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
All levels and schools pooled						
Schools that have have apparent	-0.067***	0.016	-0.000	0.039**	0.034	0.065***
random changes in gender make-up	(0.018)	(0.017)	(0.017)	(0.016)	(0.025)	(0.024)
Observations	72613	72613	72613	72613	40195	40195
All Schools	-0.064***	0.010	0.023*	0.030**	0.029*	0.042**
	(0.013)	(0.012)	(0.012)	(0.012)	(0.018)	(0.017)
Observations	144085	144085	144085	144085	80071	80071
Primary Schools						
Schools that have have apparent	-0.055***	0.005	0.027*	0.040***	0.070***	0.064***
random changes in gender make-up	(0.015)	(0.014)	(0.015)	(0.015)	(0.021)	(0.021)
Observations	65703	65703	65703	65703	33285	33285
All Schools	-0.051***	0.006	0.040***	0.031***	0.056***	0.044***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.015)	(0.015)
Observations	130507	130507	130507	130507	66493	66493
Secondary Schools						
Schools that have have apparent	-0.115*	0.053	-0.095*	0.040	-0.036	0.058
random changes in gender make-up	(0.064)	(0.062)	(0.054)	(0.054)	(0.057)	(0.055)
Observations	6910	6910	6910	6910	6910	6910
All Schools	-0.118**	0.019	-0.040	0.035	-0.028	0.032
	(0.048)	(0.048)	(0.040)	(0.041)	(0.043)	(0.042)
Observations	13578	13578	13578	13578	13578	13578
Key Stage 1						
Schools that have have apparent	-0.085***	-0.018	0.029	0.047**	N/A	N/A
random changes in gender make-up	(0.020)	(0.018)	(0.021)	(0.021)		
Observations	32418	32418	32418	32418		
All Schools	-0.076***	-0.020	0.049***	0.030**		
	(0.014)	(0.013)	(0.016)	(0.015)		
Observations	64014	64014	64014	64014		
Key Stage 2						
Schools that have have apparent	-0.022	0.029	0.023	0.032	0.070***	0.064***
random changes in gender make-up	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
Observations	33285	33285	33285	33285	33285	33285
All Schools	-0.025*	0.035**	0.025*	0.028*	0.056***	0.044***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
Observations	66493	66493	66493	66493	66493	66493
Key Stage 3						
Schools that have have apparent	-0.158*	0.079	-0.101**	-0.005	-0.085	0.008
random changes in gender make-up	(0.087)	(0.092)	(0.049)	(0.048)	(0.056)	(0.055)
Observations	4556	4556	4556	4556	4556	4556
All Schools	-0.127*	0.049	-0.035	0.033	-0.015	0.035
	(0.065)	(0.069)	(0.036)	(0.036)	(0.041)	(0.040)
Observations	8908	8908	8908	8908	8908	8908
Key Stage 4						
Schools that have have apparent	-0.083	0.033	-0.105	0.074	-0.012	0.097
random changes in gender make-up	(0.084)	(0.078)	(0.081)	(0.079)	(0.082)	(0.080)
Observations	2354	2354	2354	2354	2354	2354
All Schools	-0.112*	-0.007	-0.056	0.023	-0.055	0.017
	(0.062)	(0.058)	(0.059)	(0.058)	(0.061)	(0.060)
Observations	4670	4670	4670	4670	4670	4670

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 8 Comparison – Ordered and Random in Small and Large primary schools

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Large Primary Schools						
Schools that have have apparent random changes in gender make-up	**-0.028**	**0.031**	**0.047****	**0.051****	**0.094*****	**0.102*****
	(0.023)	(0.021)	(0.023)	(0.023)	(0.032)	(0.032)
	29502	29502	29502	29502	15456	15456
All Schools	**-0.033****	**0.006**	**0.052*****	**0.023**	**0.077*****	**0.061****
	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.024)
	58075	58075	58075	58075	30625	30625
Small Primary Schools						
Schools that have have apparent random changes in gender make-up	**-0.113*****	**-0.019**	**0.001**	**0.036**	**-0.040**	**-0.012**
	(0.031)	(0.029)	(0.031)	(0.029)	(0.046)	(0.045)
	13902	13902	13902	13902	6216	6216
All Schools	**-0.108*****	**0.015**	**0.015**	**0.059*****	**-0.032**	**0.010**
	(0.022)	(0.021)	(0.024)	(0.022)	(0.035)	(0.033)
	27338	27338	27338	27338	12320	12320
Key Stage 1 in small primary schools						
Schools that have have apparent random changes in gender make-up	**-0.151*****	**-0.013**	**0.003**	**0.063**	N/A	N/A
	(0.038)	(0.036)	(0.041)	(0.040)		
	7686	7686	7686	7686		
All Schools	**-0.146*****	**-0.002**	**0.027**	**0.073****		
	(0.028)	(0.026)	(0.031)	(0.029)		
	15018	15018	15018	15018		
Key Stage 2 in small primary schools						
Schools that have have apparent random changes in gender make-up	**-0.062**	**-0.029**	**-0.006**	**-0.005**	**-0.040**	**-0.012**
	(0.050)	(0.049)	(0.046)	(0.043)	(0.046)	(0.045)
	6216	6216	6216	6216	6216	6216
All Schools	**-0.059***	**0.037**	**-0.004**	**0.041**	**-0.032**	**0.010**
	(0.035)	(0.034)	(0.034)	(0.032)	(0.035)	(0.033)
	12320	12320	12320	12320	12320	12320

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 9 FSM Table

|	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Key Stage 1						
Proportion of the within-school cohort that are female	**-0.073***	**-0.022***	**0.052***	**0.028***	N/A	N/A
	(0.014)	(0.013)	(0.016)	(0.015)		
	[-0.25***]	[-0.07*]	[-37.64***]	[-20.10*]		
Proportion of males that receive FSM within cohort	-0.465***	0.024*	-0.390***	0.012		
	(0.015)	(0.013)	(0.016)	(0.016)		
Proportion of females that receive FSM within cohort	0.017	-0.413***	0.012	-0.360***		
	(0.014)	(0.013)	(0.015)	(0.015)		
Observations	64014	64014	64014	64014		
Adjusted R-squared	0.03	0.04	0.06	0.06		
Key Stage 2						
Proportion of the within-school cohort that are female	**-0.020**	**0.030****	**0.029****	**0.022**	**0.061***	**0.038****
	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
	[-0.07]	[0.11**]	[-0.44**]	[-0.34]	[-10.26***]	[-6.50**]
Proportion of males that receive FSM within cohort	-0.431***	-0.016	-0.371***	0.008	-0.367***	0.009
	(0.016)	(0.015)	(0.015)	(0.015)	(0.017)	(0.016)
Proportion of females that receive FSM within cohort	0.024	-0.396***	0.006	-0.374***	0.010	-0.384***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.016)	(0.016)
Observations	66493	66493	66493	66493	66493	66493
Adjusted R-squared	0.07	0.05	0.07	0.09	0.11	0.14
Key Stage 3						
Proportion of the within-school cohort that are female	**-0.116***	**0.052**	**-0.025**	**0.036**	**-0.005**	**0.038**
	(0.065)	(0.069)	(0.035)	(0.035)	(0.040)	(0.040)
	[-0.37*]	[0.17]	[0.58]	[-0.84]	[0.11]	[-0.83]
Proportion of males that receive FSM within cohort	-0.493***	-0.031	-0.461***	-0.082**	-0.472***	-0.024
	(0.064)	(0.069)	(0.037)	(0.036)	(0.041)	(0.041)
Proportion of females that receive FSM within cohort	0.052	-0.423***	0.012	-0.473***	-0.010	-0.542***
	(0.059)	(0.068)	(0.033)	(0.034)	(0.038)	(0.040)
Observations	8908	8908	8908	8908	8908	8908
Adjusted R-squared	0.04	0.02	0.12	0.15	0.21	0.25
Key Stage 4						
Proportion of the within-school cohort that are female	**-0.111***	**-0.007**	**-0.055**	**0.024**	**-0.053**	**0.018**
	(0.060)	(0.057)	(0.057)	(0.057)	(0.060)	(0.059)
	[-0.32*]	[-0.02]	[9.44]	[-4.18]	[5.56]	[-1.85]
Proportion of males that receive FSM within cohort	-0.739***	-0.043	-0.651***	-0.132**	-0.668***	-0.134**
	(0.068)	(0.062)	(0.063)	(0.060)	(0.067)	(0.066)
Proportion of females that receive FSM within cohort	0.009	-0.528***	-0.058	-0.586***	-0.045	-0.599***
	(0.063)	(0.059)	(0.060)	(0.058)	(0.064)	(0.064)
Observations	4670	4670	4670	4670	4670	4670
Adjusted R-squared	0.04	0.04	0.09	0.07	0.03	0.03

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 10 Value Added

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
All levels and schools pooled						
Proportion of the within-school cohort that are female	**-0.022**	**0.009**	**0.013**	**0.062*****	**-0.033**	**0.065***
	(0.027)	(0.027)	(0.020)	(0.022)	(0.037)	(0.035)
Observations	27270	27270	27270	27270	12962	12962
Adjusted R-squared	0.06	0.03	0.13	0.21	0.34	0.44
Pooled Secondary Schools						
Proportion of the within-school cohort that are female	**-0.106****	**-0.038**	**-0.003**	**0.076****	**-0.033**	**0.065***
	(0.041)	(0.042)	(0.026)	(0.030)	(0.037)	(0.035)
Observations	12962	12962	12962	12962	12962	12962
Adjusted R-squared	0.06	0.03	0.22	0.34	0.34	0.44
Key Stage 2						
Proportion of the within-school cohort that are female	**0.081*****	**0.066****	**0.035**	**0.045**		
	(0.031)	(0.029)	(0.032)	(0.032)		
Observations	14308	14308	14308	14308		
Adjusted R-squared	0.03	0.01	0.06	0.06		
Key Stage 3						
Proportion of the within-school cohort that are female	**-0.104**	**0.034**	**-0.017**	**0.098*****	**-0.058**	**0.088***
	(0.069)	(0.071)	(0.034)	(0.035)	(0.049)	(0.047)
Observations	8744	8744	8744	8744	8744	8744
Adjusted R-squared	0.08	0.04	0.27	0.42	0.39	0.47
Key Stage 4						
Proportion of the within-school cohort that are female	**-0.099**	**-0.174***	**0.021**	**0.009**	**-0.004**	**0.004**
	(0.097)	(0.099)	(0.047)	(0.044)	(0.057)	(0.058)
Observations	4218	4218	4218	4218	4218	4218
Adjusted R-squared	0.02	0.03	0.08	0.11	0.22	0.34
Small school Key Stage 2						
Proportion of the within-school cohort that are female	**0.071**	**0.028**	**-0.058**	**0.012**		
	(0.061)	(0.057)	(0.065)	(0.056)		
Observations	2984	2984	2984	2984		
Adjusted R-squared	0.03	0.01	0.05	0.04		
Large school Key Stage 2						
Proportion of the within-school cohort that are female	**0.081**	**0.063**	**0.046**	**0.034**		
	(0.063)	(0.053)	(0.058)	(0.061)		
Observations	5214	5214	5214	5214		
Adjusted R-squared	0.03	0.01	0.07	0.07		

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

Figure 1 The Structure of the cohorts available in the NPD and PLASC[11]

[11] From http://www.bris.ac.uk/depts/CMPO/PLUG/userguide/cohorts.pdf
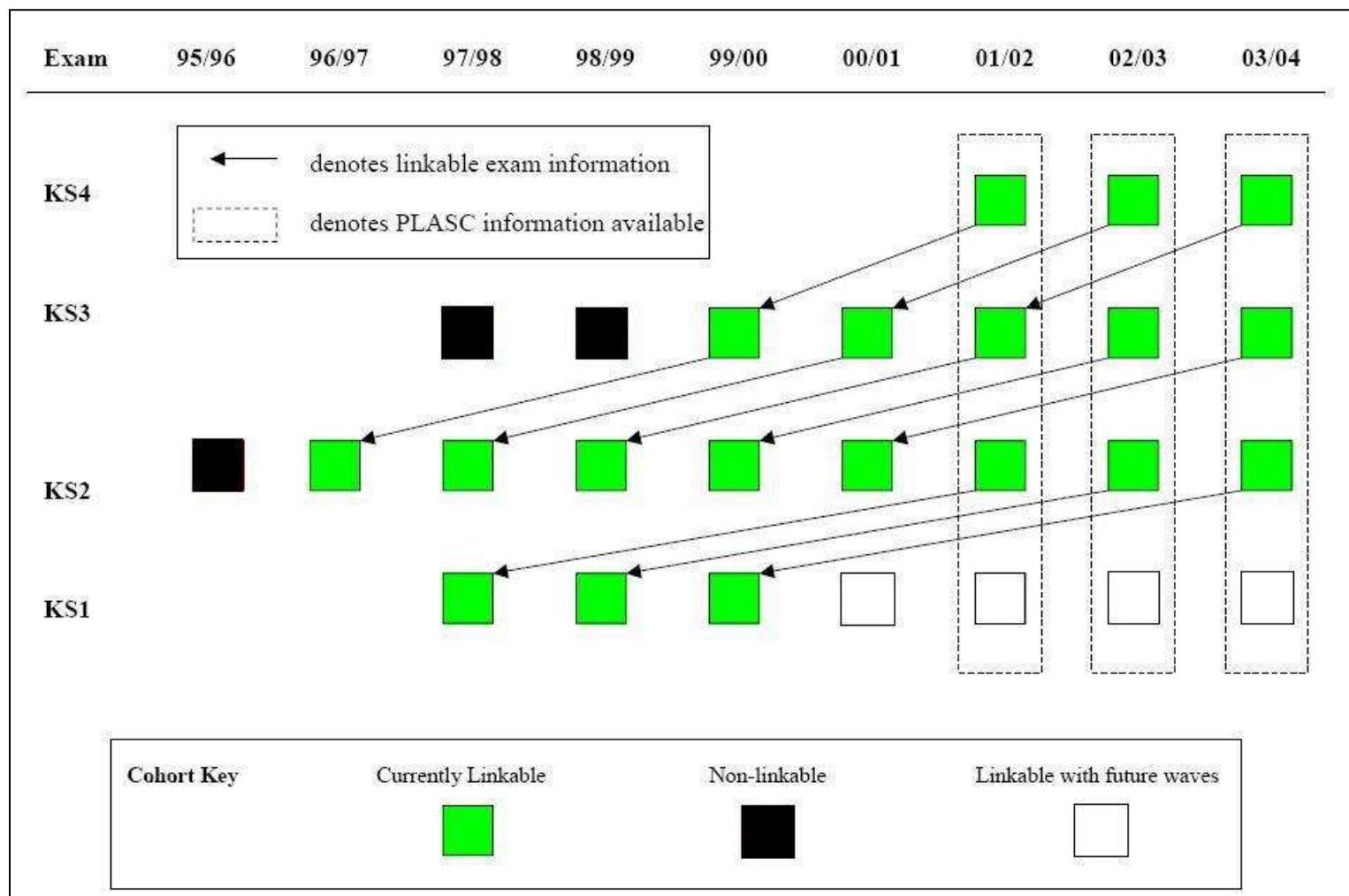
Figure 2 Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils.
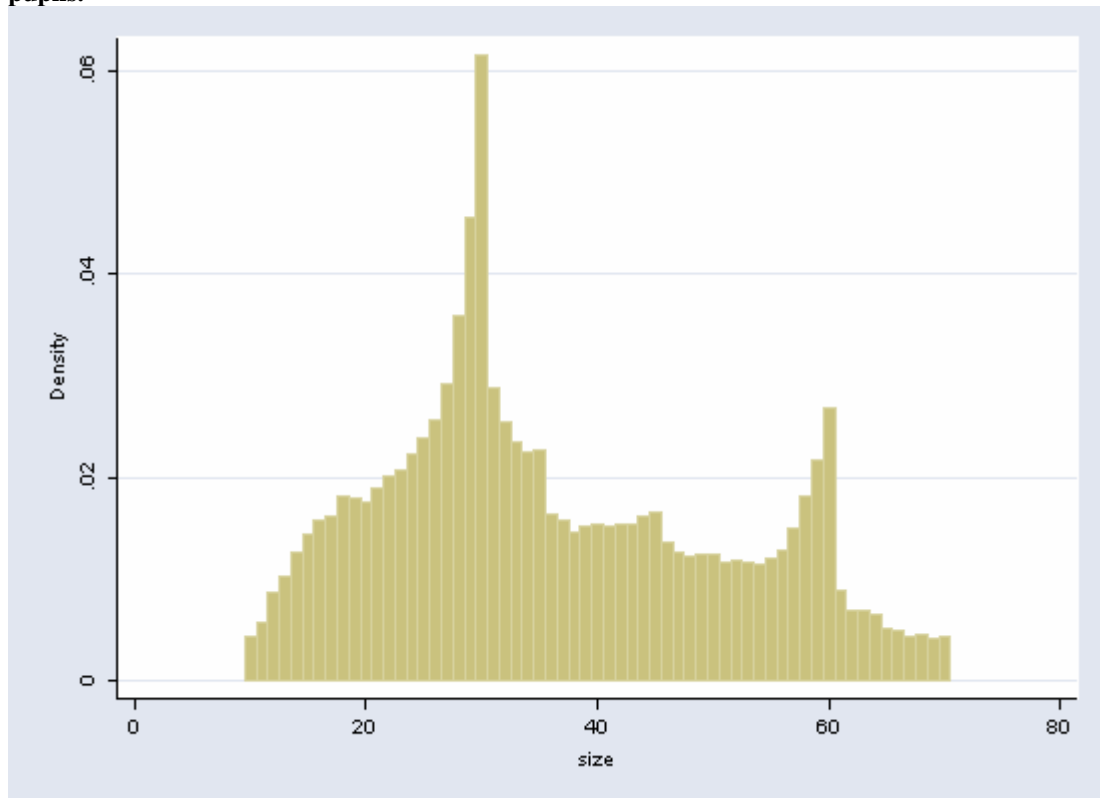

Figure 3 Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils. Pre 2002
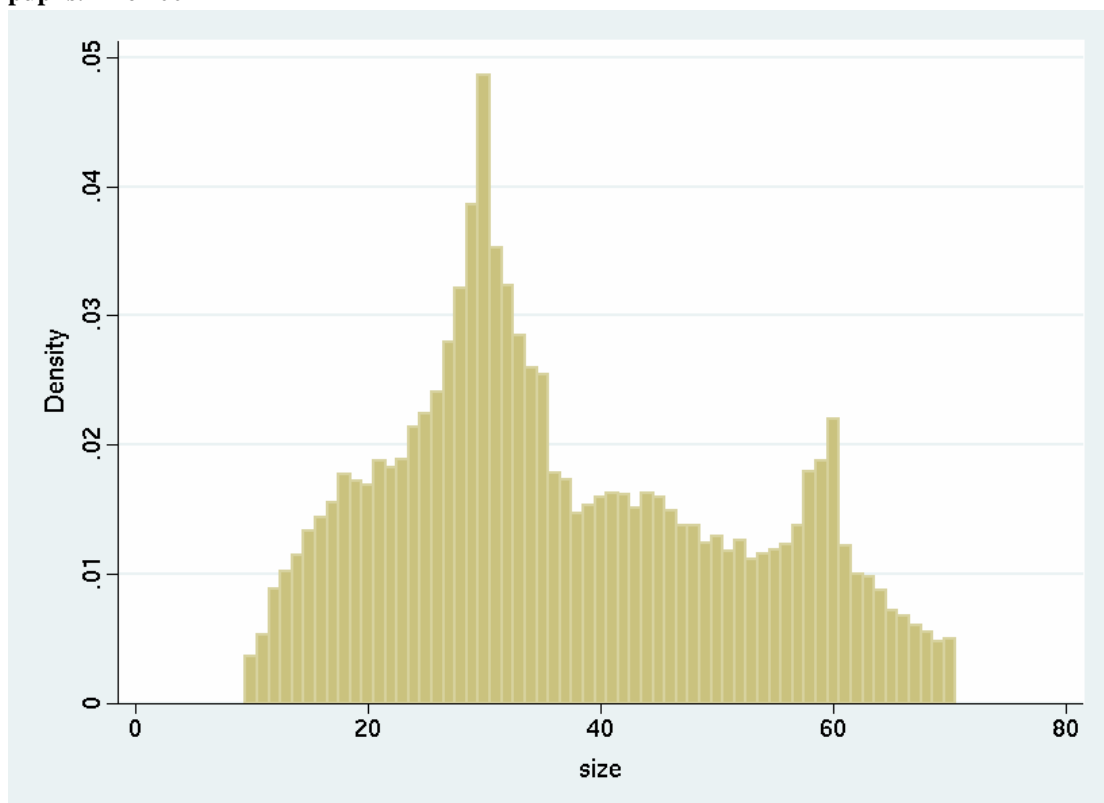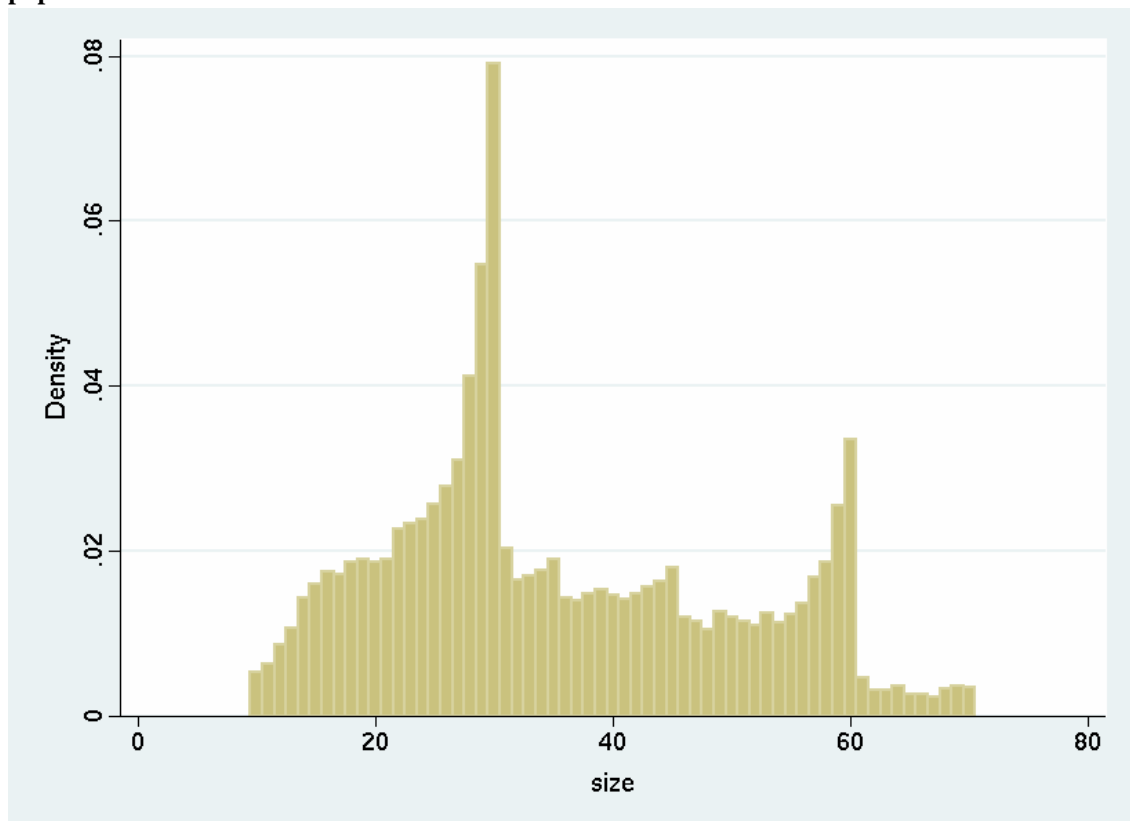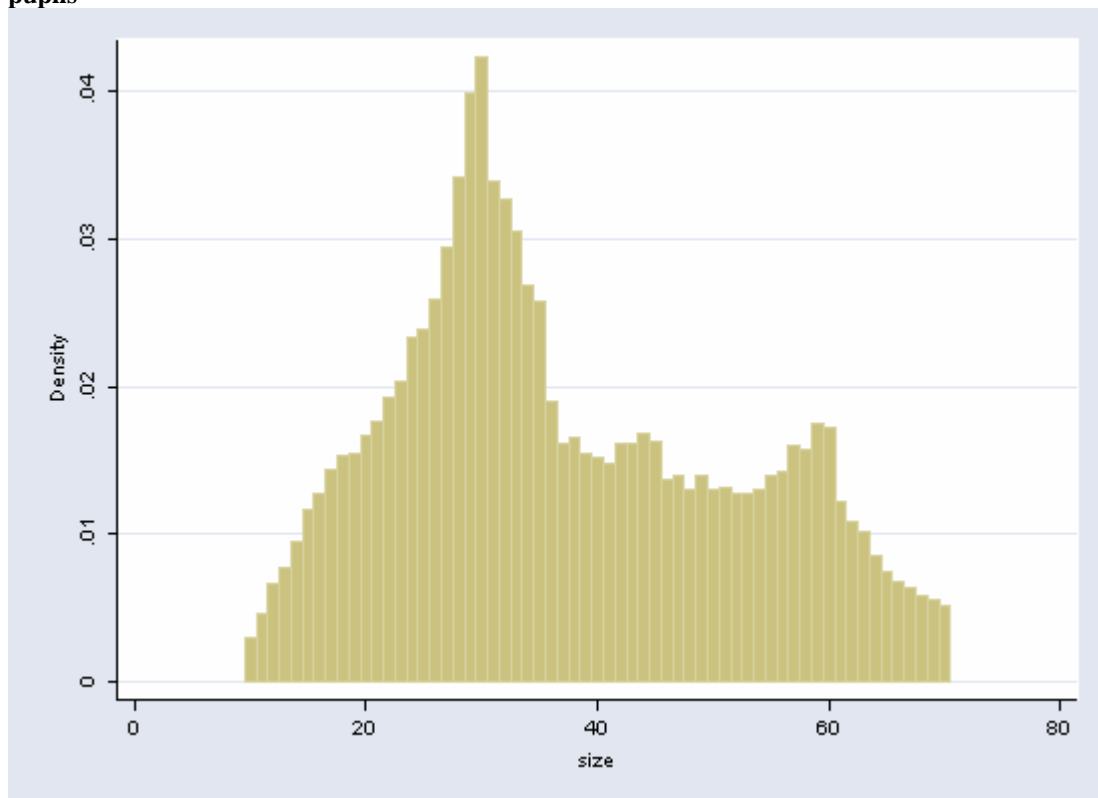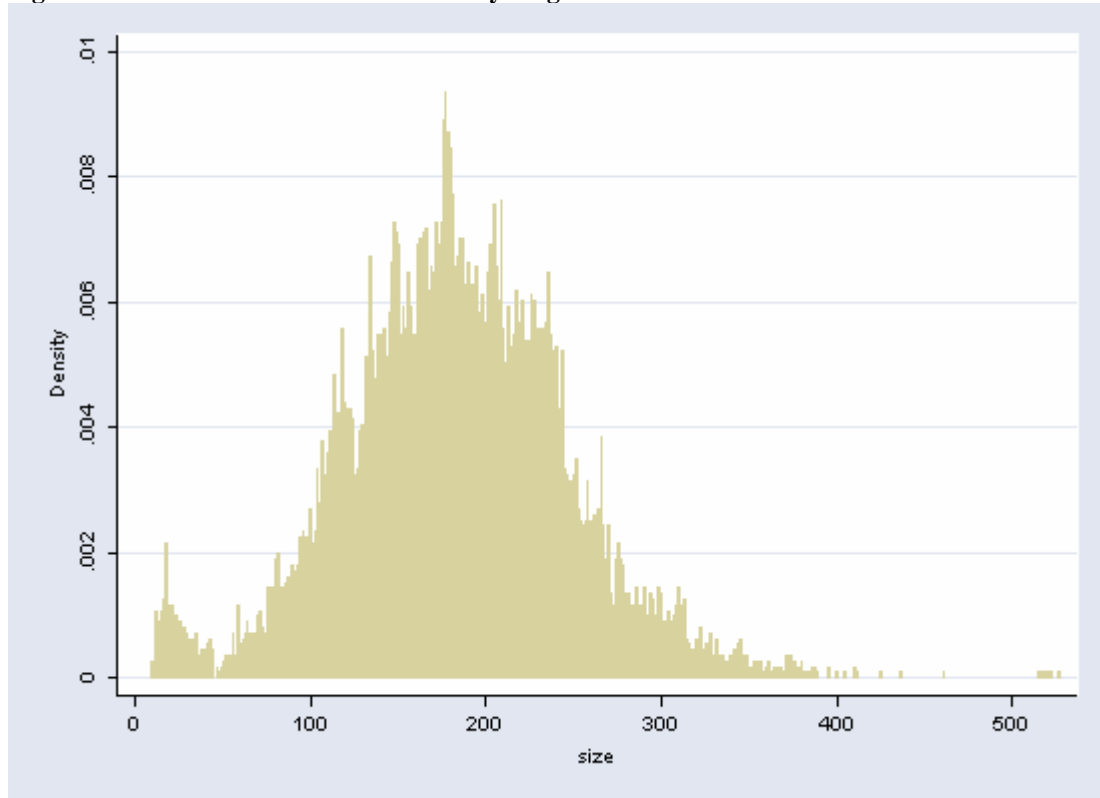
Figure 4 Distribution of School sizes at Key stage 1, focussing on schools with fewer than 70 pupils. Post 2002

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
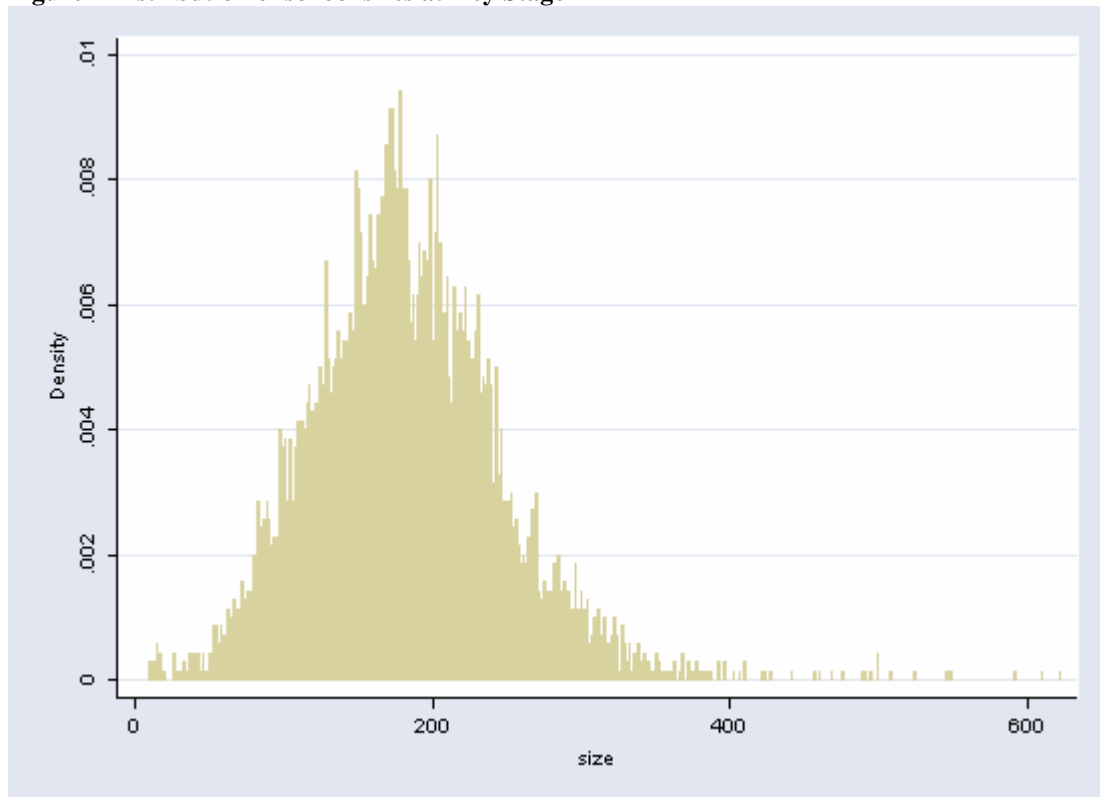
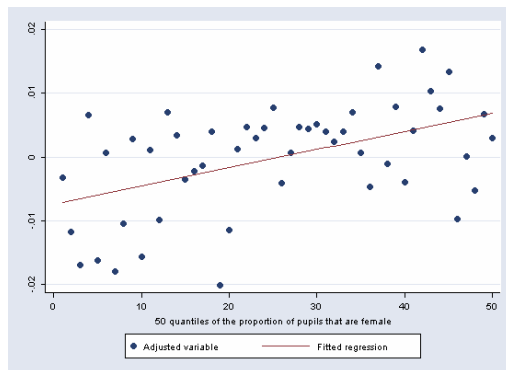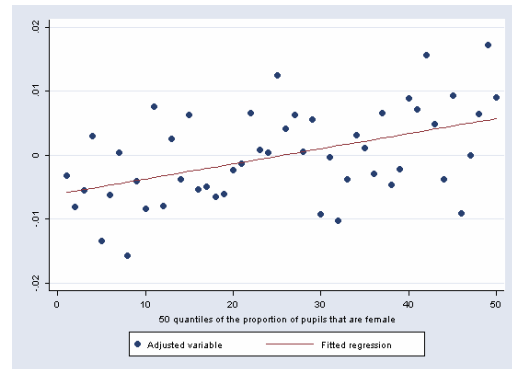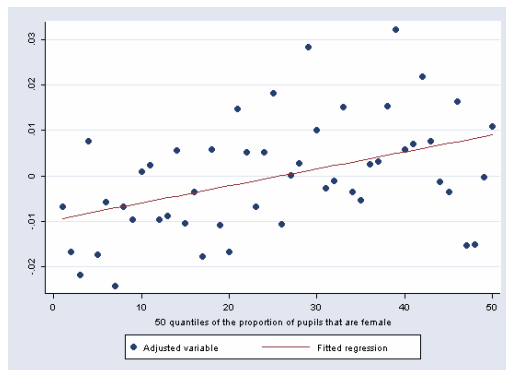**Figure 8 Adjusted variable plots of the pooled regressions.**

English

Male
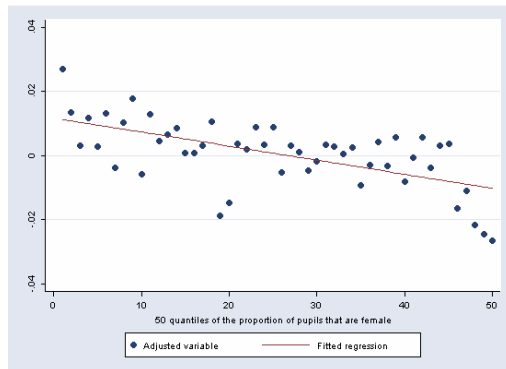
Female
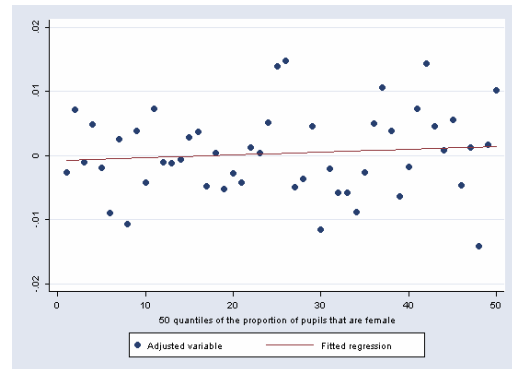
Maths

Male

Female

Science

Male

Female