The Evolution and Determinants of the Educational Gender Gap in England

Martyn Andrews
University of Manchester

Steve Bradley
Lancaster University

Dave Stott
Lancaster University

Jim Taylor
Lancaster University*

January 2006

*The authors thank The Nuffield Foundation (under grant EDU/00328/G) for financial assistance. The National Pupil Database was kindly supplied by DfES and Social Community Planning; the Youth Cohort Study is available from the UK Data Archive (`http://www.data-archive.ac.uk`). The comments from participants at various presentations are gratefully acknowledged. These include the Economics departments at the University of Brisbane, as well as the 2004 WPEG Conference at Royal Holloway.

Abstract

In this paper, we seek to explain why boys do worse than girls in examinations up to, and including, their GCSEs. Using ten sweeps of the bi-annual Youth Cohort Study (YCS), starting in 1985 and finishing in 2001, and two sweeps of the National Pupil Database (NPD) for 2002 and 2003, we define various measures of the gender gap used in the literature, and document how the gap has changed over time. Next, we explain how these gender gaps alter when controlling for personal, school, family and neighbourhood covariates, and when controlling for individual- and school-level unobserved heterogeneity. We repeat this analysis for different subjects and model how the gender gap changes at different stages in the educational process, namely at Key Stage 2 (age 11), Key Stage 3 (age 14), and GCSE (Key Stage 4, age 16).

Our overall conclusion is that the quasi-privatisation of schooling, together with the introduction of school league tables in 1992, has possibly partly contributed to a spurious observed increase in the gender gap through the 1990s, simply because the better schools have selected better girls rather boys at the margin.

Keywords: education, gender gaps, pooled cross-section and panel data
New JEL Classification: I210, I280

Address for Correspondence:

Dr. M.J. Andrews
School of Social Sciences
University of Manchester
Manchester, M13 9PL

Email: martyn.andrews@manchester.ac.uk
Phone: +44-(0)161-275-4874

Filename: ggap1.tex
Processed at: 14:34, 26th January, 2006

1 Introduction

Throughout the 1990s, the performance of girls in GCSE exams has been superior to boys. Moreover, the gap is getting wider. This issue is clearly of public concern—one only needs to read popular press each year when GCSE results are announced. In particular, there is concern for boys at the lower end of the ability distribution, whose performance has given rise to the allegation of a new culture of 'laddish behaviour'. An alternative view is that the widening gender gap does not matter if this advantage dissipates by the time the girl enters the labour market, where, of course, women fair worse than men. In many areas of gender discrimination in the labour market, the gap is getting narrower, and so one possible explanation, in the UK at least, is that the increasing education gender gap has had an impact in subsequent labour-market outcomes. It is possible that girls work harder at school knowing that they will be discriminated against later on in the labour market, an effect that has been observed for ethnic minorities by Leslie & Drinkwater (1999).

If there is a link between performance at school and subsequent labour-market outcomes, it is important to establish whether the gender gap is subject specific, and whether certain subjects are rewarded more than others in the labour market. This would seem to be the case, as there is evidence that girls do better than boys in English (and similar subjects that require a sound grasp of 'language' skills), whereas boys do better in Maths and Science. There is also evidence that there are significant wage premia for maths skills, although it refers to individuals who completed their compulsory schooling in the mid-1970s (Dolton & Vignoles 2002, Murnane, Willet & Levy 1996, Duncan & Dunifon 1998).

In this particular paper, we seek to explain why boys do worse than girls in examinations up to, and including, their GCSEs. Using ten sweeps of the bi-annual Youth Cohort Study (YCS), starting in 1985 and finishing in 2001, and two sweeps of the National Pupil Database (NPD) for 2002 and 2003, our aims and objectives are fourfold. First, we define various measures of the gender gap used in the literature, and document how the gap has changed over time. Next, we explain how these gender gaps alter when controlling for personal, school, family and neighbourhood covariates, and when controlling for individual- and school-level unobserved heterogeneity. Third, we repeat this analysis for different subjects. Finally, we model how the gender gap changes at different stages in the educational process, namely at Key Stage 2 (age 11), Key Stage 3 (age 14), and GCSE (Key Stage 4, age 16).[1]

[1] All of these examinations have nationally agreed standards and are taken by all pupils in England and Wales.

In Section 2 we review the literature. In Section 3 we outline the methodology we adopt in this paper, before describing the data we use in Section 4. Various measures of educational attainment have been discussed in the literature; the ones we use for our own empirical work are discussed in Section 5. Whether one should use absolute or relative gender gaps is discussed in Section 6. Our results are discussed in Section 7. Section 8 concludes.

2 Literature review

There are two broad strands of literature concerned with the education gender gap (hereafter the gender gap). One strand is from educational research, where the majority of the previous work has focused either on the measurement and description of the evolution of the gender gap, or the qualitative analysis of the determinants of the gender gap. There are also a small number of quantitative studies, arising from the school effectiveness literature, that use multi-level modelling techniques to investigate the determinants of the gender gap. However, a more recent strand of literature has emanated from education economists who adopt a production function framework and econometric techniques to investigate the causes of the gender gap.

2.1 Educational research

An example of a descriptive study of the gender gap is by Gorard, Rees & Salisbury (1999), who use individual level data for 1997 for Wales. Their findings show that the gender gap does not increase between Key Stage 3 and Key Stage 4, and that the gap itself is not uniform over the attainment distribution. Rather the gender gap is greatest at the upper end of the attainment distribution (i.e. for Key Stage 1 through to Key Stage 4 or GCSE). They also show that the gender gap is greatest in English and Welsh, whereas there is no difference between the attainment of boys and girls in Maths and Science. The limitation of this kind of study is that it does not offer an explanation for these findings. Another limitation is the way in which they measure the education gap, focusing on relative rather than absolute differences in educational attainment.

School effectiveness studies, such as Wong, Lam & Ho (2002), use multi-level modelling techniques to analyse the gender gap. They analyse the gender gap for Hong Kong in 1997, a former British colony which retains many of the features of the British education system. Girls get better exam results when they are educated in single-sex schools whereas for boys mixed schools are better. In contrast to the findings for Wales, they also show that girls did better in English and Maths. Using similar techniques, Thomas, Sammons, Mortimore & Rees (1997) and Mortimore & Sammons (1994) also find for Britain that school type does affect the gender gap.

The bulk of educational research on the gender gap is qualitative, based mainly on case studies of pupils and/or schools. This literature also tends to be prescriptive, suggesting that the gender gap is caused by a particular feature of schools or pupils, in spite of the difficulty in generalising their results. Salisbury, Rees & Gorard (1999) and OFSTED (2003) provide very good reviews of this literature, and its findings on the 'causes' of the gender gap can be classified into five categories.

The first set of causes relate to the organisation of teaching and learning, curriculum design and assessment. It is argued that girls have better literacy skills when they enter secondary school, and that current methods of teaching favour the learning style of girls. Girls prefer narrative forms of assessment, whereas boys prefer multi-choice questions, for instance, and the current GCSE examination system is biased in favour of a narrative approach. Interestingly, qualitative researchers have found no compelling evidence that girls benefit more than boys from the focus on coursework in the GCSE curriculum (Murphy 1999). What has been identified, however, is an inequality in tiering practices and misrepresentation of student achievement which benefits girls (Elwood & Murphy 2002). Thus, girls are more likely to be entered for the 'higher' tier exam, which means that they have the opportunity to achieve higher grades. School ethos is often cited as a factor that has changed over time and that this has mitigated against the educational attainment of boys. Specifically, it is argued that boys prefer discipline, expectations of high standards and a promotion of a sense of commitment, which are missing in many of Britain's schools today.

A second set of causes of the gender gap refer to 'school organisation'. School setting regimes, determined by tests in year 7 (age 11), lock boys into a pattern of under-achievement. If boys are placed in lower sets, then they are more likely to have poorer peer groups, which reinforces their under-achievement. Furthermore, the decision to tier pupils' exam entries and to stream pupils is reinforced by competition between schools in the current quasi-market environment. This behaviour arises from the annual publication of league tables, which leads to target setting in schools to maximise the proportion of pupils obtaining 5 or more GCSE grades A*-C.

A third set of causes is referred to as 'the culture of laddishness'. Boys have worse 'non-cognitive' skills, including the inability to pay attention in class, to work with others, to organise and keep track of homework or class materials and to seek help from others (Jacob 2002). Furthermore, academic work is susceptible to peer group influences, especially for boys. Warrington, Younger & Williams (2000) and Tinklin (2003) argue that girls can work hard and still be part of the 'in crowd', whereas the motivation of boys towards academic work falls from year 8 onwards (Warrington & Younger 1999). This has given rise to the claim that boys are 'disaffected', 'disappointed', 'disappeared' — truants and exclusions from school.

The two remaining sets of causes of the gender gap can be referred to as school effects and poverty. In terms of school effects OFSTED (2003) argue that 'Contemporary inspection evidence showed that about one secondary school in five was weak in meeting the particular needs of one or other sex.' (p7). This implies a school fixed effect on educational attainment, over and above the causes discussed above, however it is not clear whether this effect favours girls more than boys. The effect of home background on educational attainment has been very well documented, but little evidence exists on the impact of home background on the gender gap. It is suggested by OFSTED (2003) that a poor home background reduces the motivation of boys more than for girls, hence leading to inferior exam performance.

2.2 Education Economics

Very little work has been conducted by economists on the gender gap. Dolton, Makepeace, Hutton & Audas (1999), using YCS1–7, investigate the impact of the replacement of the GCE system by the GCSE system. (Recall the points made earlier about the nature of the GCSE assessment.) Boys and girls improved their performance, but the effect for girls was much larger — the probability of obtaining 5 or more GCSEs graded A-C increased by 1.6 percentage points for boys versus 6.4pp for girls. Asian girls have lower attainment, possibly because of cultural factors, as do girls with a larger number of siblings. Girls with a father in intermediate or professional occupations and those from a single parent family do better. Recently, Atkinson & Wilson (2003) have used the National Pupil Database (NPD) for pupils who sat their GCSE exams in 1999 and their Key Stage 3 tests in 1997. They provide evidence for a widening of the gender gap between age 14 and 16. In fact, by age 16 girls are better by the equivalent of one grade C, and they perform better in English, Maths and Science, though the latter are small. In *separate* regressions for boys and girls they also find evidence of school effects — grammar and single-sex schools raise attainment — but these effects disappear in value added models. Poverty reduces attainment. Perhaps the most rigorous study of the gender gap is Burgess, McConnell, Propper & Wilson (2003) who use the NPD for 2001 and show a consistent gender gap across the attainment distribution driven mainly by girls' superior performance in English. The gap is not affected by school factors but it is positively affected by school performance. The gender gap is also negatively affected by poverty and the proportion of boys within a cohort. One problem with this study is that it might be misspecified by including school performance and school factors, since the former is a function of the latter.

Previous work on the gender gap is dominated by educational research, and of the three types of research in this area, qualitative studies dominate. Very little research has been conducted by economists. There are some conflicting findings, however, previous research suggests a wide range of potential causes of the gender gap. Unfortunately, some of these causal variables are either unmeasurable (e.g. teaching and learning styles, pupil attitudes to school or motivation) or unobservable in national datasets (e.g. tiering practices, school ethos). This suggests that it is important to control for individual- and school-level unobserved heterogeneity, which we do in this paper. Moreover, unlike previous research, we have data from 1985-2003, which enables us to obtain consistent estimates of time-series movements in the gender gap, and to identify structural breaks in the gap due to changes in the system of examinations. Finally, by combining the YCS and NPD data with School Performance and School Census data, we have a wider range of covariates, including many school and family variables.

3 Methodology

The methodology we adopt in this paper is to estimate different specifications of the so-called education production function. In words, this means modelling (various measures of) the educational attainment y_i for a sample of pupils i as a function of a girl dummy g_i and various observed covariates. Our aim is to estimate the education gender gap, which is the parameter associated with g_i unless there are additional interactions with the covariates. The covariates can be classified into fixed characteristics $\mathbf{x}_i$ (e.g. family background), school-level variables $\mathbf{x}_{s(i)}$ (e.g. school size, gender mix), and neighbourhood variables $\mathbf{x}_{r(i)}$. In other words, s indexes school and r indexes neighbourhood. It is usual to think of $\mathbf{x}_i$ as being time-invariant, whereas school-level and neighbourhood variables can vary by time in pooled cross-sections or panel data. This is true even if the pupil does not change school or neighbourhood.

One issue is whether one should even bother with adding the observed covariates, as it is possible to argue that g_i cannot be correlated with any observables: girls are no more or less likely to be born to certain families from certain backgrounds than boys. The extreme version of this view also argues that pupils (or their parents) have had little opportunity to make real choices about their lives before taking their GCSEs, and so the selection issue disappears. On the other hand, it might be the case that school choice is important, as is the choice of subjects taken at GCSE. There is much evidence that suggests that creating a quasi-market in school choice throughout the 1990s has affected examination performance (Bradley, Crouchley, Millington & Taylor 2000, Bradley & Taylor 2004). Boys tend to take more science/maths type subjects, which is why we estimate our models disaggregated by subject k as well as examine aggregate measures. The point here is that significant interaction of g_i with covariates, or correlations of g_i with covariates, are more likely with $\mathbf{x}_{s(i)}$ than with $\mathbf{x}_i$ or $\mathbf{x}_{r(i)}$. Accordingly, we have a large range of school-level covariates mapped into the NPD and YCS from other sources. The extent to which gender is correlated with any of the observed covariates determines the extent to which raw and conditional differentials diverge.

Even if g_i is uncorrelated with many observable covariates, it is almost certain that g_i is correlated with unobservables. We identify two specific unobserved heterogeneity terms: the unobserved innate ability of individuals α_i (including attitudes to education and motivation) and the unobserved school quality η_s (e.g. ethos, discipline, etc.). Given the conclusions from the literature survey—especially some of the qualitative findings—we ask whether α_i is positively correlated with g_i, ie are girls cleverer, more motivated, better at doing exams, than boys? Similarly, is η_s positively correlated with g_i, ie do better (mixed) schools attract girls rather than boys? Finally, we ask whether better pupils are more likely to attend better schools, ie is α_i positively correlated with η_s?

Controlling for these three correlations is very important for the usual omitted-variables-bias type arguments. If one can obtain useful estimates of these correlations, that is even better. If it is the case that g_i is uncorrelated with the observed covariates $\mathbf{x}$, then if one ignores the issue of unobserved heterogeneity, one is simply left with documenting the raw differentials reported in the press or available on the DfES website. It is focusing on unobserved heterogeneity that represents value-added in this particular study (but see Burgess et al. (2003)).

4 Datasets

We use two different datasets in this paper, namely ten sweeps of the Youth Cohort Study (YCS) and two sweeps of the recently released National Pupil Database (NPD). These are used to form a long run of pooled cross-section data and a much shorter panel dataset. The two key differences are that exam performance for each individual is observed three times in the latter, and that the latter records the population. For both datasets, we match in school-level information from the annual School Performance Tables and the School Census, both provided by the DfES, and neighbourhood information from NOMIS.

4.1 Pooled cross-section data

Here we pool ten sweeps of the bi-annual YCS, starting in 1985 and finishing in 2001 (YCS2-11), and append two sweeps of the NPD for 2002 and 2003 (NPD1 and NPD2). The dependent variable y_i comprises various measures of exam performance at GCSE from YCS4 onwards and GCE in YCS2 and YCS3. (The change from GCE to GCSE took place in 1987.) The key feature of this pooled cross-section data is that we are able to observe the time-series movements in the gender gap since 1985 on a consistent basis. The personal and family covariates $\mathbf{x}_i$ comprise gender g_i, age, ethnicity, parental occupation, household status, whether a single parent and housing tenure. The school-level covariates $\mathbf{x}_s$ comprise information on teachers, pupil concentration, pupil-teacher ratios, school-size, the composition of the school and measures of the pupils' peer group. Neighbourhood variables $\mathbf{x}_r$ include the local unemployment rate, and the proportion of professional and managerial workers in the locality. However, certain important variables are not observed before YCS6, including whether the school is single sex. Also, we do not have consistent school identifier s for YCS2-5, but do from YCS6 onwards.

Because this is a pooled cross-section dataset, we cannot control for individual heterogeneity α_i, but can control for school-level heterogeneity η_s. (This is slightly compromised by not having a consistent measure for s for YCS2-5.)

4.2 National Pupil Database 2002 & 2003

By contrast, not only is the NPD a panel dataset, with examinations at Key Stage 2, Key Stage 3, and Key Stage 4, it is the population of pupils taking their exams in England in state-funded schools.[2] We use two cohorts, for pupils who took their GCSEs in 2002 and in 2003. Table 1 illustrates.

The observed covariates are similar to those in the YCS. Note that age is measured to the month. It is well-established that exam performance improves with age within a given school year—we need to establish whether this varies with gender. YCS has much better information on family background (for example, fathers' and mothers' background) whereas only eligibility for free school meals is available in the NPD, a proxy for family income.

The advantage of the NPD panel, compared with the pooled cross-section data, is that we observe educational attainment more than once, and therefore we can control for and estimate α_i as well as η_s. Specifically, we can model how the gender gap changes between ages 11 and 14, and between 14 and 16, but we *cannot* estimate its level because gender is a fixed-effect. Briefly, we regress the change in exam performance between KS2 and KS3 on the change in observables, pooled together with the change in exam performance between KS3 and KS4 on the corresponding change in observables. Making sure that educational attainment at KS3 and KS4 are comparable is an important issue here. We also examine the extent to which α_i, η_s and g are correlated with each other.

Because very few individual-level variables change over time, identification for these fixed-effects techniques relies heavily on school and neighbourhood covariates. These do vary over time, so identification does not need individuals to change school or neighbourhood (approximately 8,000 in each of NPD1 and NPD2) unless we want to control for α_i and η_s simultaneously.

5 Measures of educational attainment

As mentioned already, there are a number of ways of measuring exam performance. These include:

- Pass/fail for subject k (A*-C GCSE)
- Number of Passes (number of A*-C GCSEs)
- Binary 5+ Passes (5+ A*-C GCSEs)
- Binary 8+ Passes (8+ A*-C GCSEs)

[2] Key Stage 4 and GCSE are used as synonyms throughout.

- Points Score

Binary 5+ Passes is the headline measure discussed almost exclusively in the press. Note that pupils can take up to 15 subjects. A pass is Grade A*-C. In the NPD, at Key Stage 2 and Key Stage 3, exams are only taken in English, Maths and Science. These different measures are now discussed in some detail.

5.1 Number of Passes

This simply counts the number of passes (number of A*-C GCSEs) for each individual. In Figure 1, we plot histograms for YCS3 (1986), YCS6 (1991), YCS10 (1999) and NPD1 (2002), to show how this variable has changed over time.[3] Superimposed is the theoretical Poisson distribution, computed with the corresponding sample mean (also shown).

A number of observations are noteworthy. First, girls and boys have identical distributions in YCS3. Second, both distributions move to the right through time (attainment increasing), but the girls' distribution moves to the right faster than the boys', ie the gender gap is increasing. The distribution has two modes and is therefore clearly not Poisson distributed, with right-most mode at 8/9/10 passes getting bigger through time.

The YCS is a non-random sample. The proportion of girls in the sample varies between 53% and 55%, whereas it is 49.5% in the population. It is generally thought that it is low-achieving boys who tend towards non-response. Given that the NPD is the population, comparing YCS10 and NPD1 is very informative, albeit they are three years apart.[4]

The average number of GCSE passes in YCS10 looks as if it is too high because zeros are under-represented compared with population NPD1; otherwise the distributions of NPD1 and YCS10 look very similar. This non-response is not a problem if it affects boys/girls equally. However, the proportion of boys with zero GCSEs is lower than that of girls (11.2% girls compared with 17.6% boys), and are different from the NPD1 three years later: 19.1% girls compared with 28.7% boys) This, in turn, suggests that there might be a structural break in the time-series between YSC10 and NPD1 for the raw gender gap, although the actual gaps suggest that

[3] A complete set of figures are available on request.

[4] Ideally one would want to compare NPD1 a with YCS11, but the number of pupils with zero passes looks implausibly low in the latter. We are examining whether this is genuine.

this is a small problem. For YCS10, the raw gender gap is 6.12-5.29=0.83; for NPD1 it is 5.18-4.17=1.01. In short, the means are different comparing 1999 with 2002, but the gaps are (probably) not.

5.2 Binary 5+ Passes

This is a binary variable, defined as $1(y_i \geq 5)$ where y_i is the number of A*-C GCSEs (as above) and 1( ) is the indicator function. Hence $\bar{y}$ for this measure is the proportion of individuals who have 5+ A*-C GCSEs. This is the "headline" statistic quoted in all popular analyses of gender gaps, and also in a lot of academic studies. In Figure 1 above, the reader could imagine a vertical line located between 4 and 5 passes, and count the proportion located to the right of it. As this is not easy to visualise, the reader is asked to wait until our discussion of Figure 3 below. An analogous measure for 8 passes is also analysed, to see whether different conclusions emerge at the top end of the exam performance distribution.

5.3 Points Score

This (final) measure sums a 'score' q for each grade (Grades 'A*' or 'A' scores $q = 7, \ldots,$ Grade 'G' scores $q = 1$) over all GCSEs taken by each individual. These are plotted in Figure 2. Most of the discussion in Section5.1 applies here: girls and boys have similar (but not identical) distributions in YCS3, and both distributions move rightwards, with the girls' distribution moving to the right faster than boys. However, here the distribution looks fairly close to being Normally distributed, except for a drop in the distribution at 70 points (some schools operate a maximum of 10 subjects at 7 per A/A*), and for some censoring at zero points score, especially in earlier years.

For this Points Score measure, the left-most part of the distribution (ie those with zero score) are over- (not under-) represented in YCS10 compared with NPD1. Why individuals who get no points whatsoever are more likely to end up in the YCS is not clear to us, although these are a much smaller proportion than those who fail all their GCSEs. Apart from this small degree of censoring, NPD1 and YCS10 look very similar (again). Again there is a possible structural break in the gender gap, being $45.1 - 41.3 = 3.8$ for YCS10 and $40.7 - 36.0 = 4.7$ for NPD1. We cannot tell, until YCS11/12 become available, whether this differential of one point (one grade per pupil) is a real structural break between YCS and NPD, or whether it actually occurred in the three-year period.

5.4 Summary

Although a count variable, the Number of Passes measure (number of A*-C GCSEs) does *not* look Poisson distributed. This suggests using standard count data techniques is inappropriate, which is why we choose not to model this particular measure in the econometrics. Although we could possibly use a non-parametric technique that compares boys/girls at each count, with common parameter across all counts, we need to choose techniques that lend themselves easily to handling fixed-effects.

For both Number of Passes (Figure 1) and Points Score (Figure 2), the distributions for boys and girls are very similar, which means we can restrict our attention to only comparing means between the two groups. This also means that comparing the cumulative frequencies of the Number of Passes distributions (ie comparing the proportions who get 5+ A*-C GCSEs) is going to give the same conclusions as comparing the sample means, as 5+ passes is located fairly close to the median of the two distributions. For this reason, and because Binary 5+ Passes is the standard measure in the literature, we use this measure in the econometrics below.

The Points Score distribution looks very much like a censored Normal (Tobit), so we also use this in the econometrics below. Controlling for unobserved heterogeneity in a Censored Normal distribution is not at all straightforward, but as the degree of censoring is very small, we are confident that using OLS is appropriate.

To conclude, we model 3 dependent variables throughout:

- Binary 5+ Passes
- Binary 8+ Passes
- Points Score (unlogged)

In pooled regressions, these are estimated by Logit and OLS respectively. Using the LPM instead of the Logit is often a very useful approximation, especially when we come to control for unobserved heterogeneity. Tables of descriptive statistics for these three variables are given in Tables 2, 3 and 4.

6 Absolute versus relative gender gaps

There is a debate in the literature, especially amongst educationalists, as to whether one should use absolute or proportional (or relative) differentials. Because the former is the one quoted almost exclusively in public debate, educationalists, who favour the latter, label this the 'politician's error'.

The absolute raw differential is given by:

$$\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0)$$

and is estimated by $\widehat{\beta}_1 = \bar{y}_1 - \bar{y}_2$ in the regression

$$y = \beta_0 + \beta_1 g + u$$

where $\bar{y}_1$ is mean attainment for girls, and $\bar{y}_2$ is mean attainment for boys.

The proportional raw differential is given by:

$$\frac{\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0)}{\mathrm{E}(y \mid g = 0)}$$

and is approximated by

$$\mathrm{E}(\log y \mid g = 1) - \mathrm{E}(\log y \mid g = 0)$$

under certain conditions. This suggests taking logs of the dependent variable and estimating

$$\log y = \beta_0 + \beta_1 g + u.$$

The obvious estimator of the proportional gap is $\bar{y}_1/\bar{y}_2 - 1$.

Figure 3 shows that this is a very important issue (see also Table 2). As both girls and boys have improved their exam performance throughout the 1990s, using the 5+ Binary Passes measure, the absolute gap has got bigger almost every year since the introduction of GCSEs in 1987 (except 1991) whereas the proportional gap increased very quickly until 1990, then flattened off, and has fallen every year since 1999. Using the proportional gap would suggest that there isn't a problem at all. It is clear that girls' level of attainment is higher *and* is growing at faster rate. It is easy to show algebraically that, when girls' level of attainment is higher and growing at a faster rate than boys, it is possible for the relative gap to falls. As this makes no sense, it is simply incorrect to use the latter.

Some educationalists (Gorard, Rees & Salisbury 2001, p.128) use the following proportional measure:

$$\frac{\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0)}{\mathrm{E}(y \mid g = 1) + \mathrm{E}(y \mid g = 0)}$$

To see how this relates to the more usual proportional measure discussed above, define the following proportionate differential

$$\frac{\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0)}{\mathrm{E}(y)} = \frac{\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0)}{p\mathrm{E}(y \mid g = 1) + (1 - p)\mathrm{E}(y \mid g = 0)}$$

where $p \equiv \mathrm{E}(g)$. If $p = 0$ this becomes the proportionate measure above; if $p = 1$, we get a proportionate measure normalised on girls rather than boys. Since the proportion of girls in the population is one-half, setting $p = 1/2$ is also sensible. This, however, is double the Gorard et al. measure.

In the rest of this paper, we ignore proportionate measures. Thus, for Points Score, we use unlogged y. For Binary 5+ Passes,

$$\mathrm{E}(y \mid g) = \Lambda(\beta_0 + \beta_1 g)$$

where Λ is the Logistic distribution function, and so

$$\mathrm{E}(y \mid g = 1) - \mathrm{E}(y \mid g = 0) = \Lambda(\beta_0 + \beta_1) - \Lambda(\beta_0)$$

This "Logistic differential" is estimated as $\bar{y}_1 - \bar{y}_2$. Note that adding covariates to the Logit regression means computing[5]

$$\Lambda(\beta_0 + \beta_1 + \mathbf{x}\boldsymbol{\beta}) - \Lambda(\beta_0 + \mathbf{x}\boldsymbol{\beta}).$$

A final issue discussed in the literature is whether one should normalise on the number of GCSEs each individual is entered for, denoted z_i in Table 2. Our view is that this disadvantages girls, who always take more GCSEs than boys. It is easier to pass five exams if only entered for eight rather than ten, although the gap in $\bar{z}$, in fact, averages about one-third of a GCSE.

[5] We should use Stata's `logit`, followed by `mfx`. This is very time-consuming, and so we use `dlogit2`, even though it treats all variables as continuous. It makes virtually no difference to the results. Also note that the LPM is useful in certain circumstances.

7 Results

7.1 Pooled cross section methods

In this subsection, we report results using the pooled (YCS/NPD) cross section dataset. The basic model, for each cohort t, is:

$$y_i = \gamma g_i + \mathbf{w}_i \boldsymbol{\beta} + u_i.$$

The specification of $\mathbf{w}_i$ varies as follows:

$$\begin{aligned}
&\mathbf{w} = [1] \qquad \text{if "raw"} \\
&\mathbf{w} = [1, S, Q, gS, G] \qquad \text{if "base"} \\
&\mathbf{w} = [1, S, Q, gS, G, \mathbf{x}] \qquad \text{if "full"}
\end{aligned}$$

where S is a selective school (that is, operates some kind of selection policy, usually with an entry exam), Q denotes a single-sex school, G denotes an all girls' school, (and $B \equiv Q - G$ defines an all-boys' school). As above, $\mathbf{x}$ denotes all the other individual, school and neighbourhood covariates, including month-of-birth dummies (all cohorts except YCS8-10). The issue here is whether any of the additional covariates are correlated with g, because, if not, they can be dropped from the analysis. This would leave us with the Base Model, whose virtue is its simplicity. Notice that the Base Model has two main effects, S and Q, and their corresponding interactions, as G can be thought of as equivalent to gQ.

Separate regressions are run for each cohort. It is possible that girls and boys respond differently, in terms of their educational attainment, to the same sets of circumstances, ie family background and school. We have carefully checked for all possible interactions between g and all the covariates, and find that the *only* interactions are the two in the Base Model.

Thus we write

$$y = \gamma_2 + (\gamma_1 - \gamma_2)g + \beta_2 S + (\beta_1 - \beta_2)gS + \beta_4 Q + (\beta_3 - \beta_4)G + \mathbf{x}\boldsymbol{\beta} + u \tag{1}$$

It is our view that the appropriate definition of the gender differential is not at all obvious. Suppose we start with a girl and, in a thought experiment, imagine what would happen if she became a boy. The problem arises in what one assumes happens to this pupil in terms of schooling, since, in this particular thought experiment, one cannot necessarily leave this pupil at the same school. We think that the most sensible assumption is that, if she starts at an all-girls' school, the pupil moves to an all-boys' school. Then we have the following four differentials $\mathrm{E}(y_i \mid g_i = 1) - \mathrm{E}(y_i \mid g_i = 0)$:

g	S	Q	differential
1	0	0	$\gamma_1 - \gamma_2$
1	1	0	$\gamma_1 - \gamma_2 + \beta_1 - \beta_2$
1	0	1	$\gamma_1 - \gamma_2 + \beta_3 - \beta_4$
1	1	1	$\gamma_1 - \gamma_2 + \beta_1 - \beta_2 + \beta_3 - \beta_4$

Thus the gender gap for any girl i (who is now a boy) is given by

$$\mathrm{E}(y_i \mid g_i = 1) - \mathrm{E}(y_i \mid g_i = 0) = (\gamma_1 - \gamma_2) + (\beta_1 - \beta_2)1(S_i = 1) + (\beta_3 - \beta_4)1(Q_i = 1).$$

Averaging over all the girls in the sample, and using estimates:

$$\widehat{\Delta}_1^* = (\widehat{\gamma}_1 - \widehat{\gamma}_2) + (\widehat{\beta}_1 - \widehat{\beta}_2)\bar{S}_1 + (\widehat{\beta}_3 - \widehat{\beta}_4)\bar{Q}_1, \tag{2}$$

where $\bar{S}_1$ is the proportion of girls who go to selective schools and $\bar{Q}_1$ is the proportion of girls who go to single-sex schools. Repeating the argument, but starting with boys, gives

$$\widehat{\Delta}_2^* = (\widehat{\gamma}_1 - \widehat{\gamma}_2) + (\widehat{\beta}_1 - \widehat{\beta}_2)\bar{S}_2 + (\widehat{\beta}_3 - \widehat{\beta}_4)\bar{Q}_2,$$

where $\bar{S}_2$ is the proportion of boys who go to selective schools and $\bar{Q}_2$ is the proportion of boys who go to single-sex schools. In practice, it makes very little difference which is used. We use $\widehat{\Delta}_1^*$ throughout.

Notice that $\mathbf{x}$ does not contribute to the conditional differential, even though it might be correlated with g. It is only the interaction terms gS and G that contribute, in addition to g itself.

For each cohort t, it is also true that

$$\bar{y}_1 = \widehat{\gamma}_1 + \widehat{\beta}_1\bar{S}_1 + \widehat{\beta}_3\bar{Q}_1 + \bar{\mathbf{x}}_1\widehat{\boldsymbol{\beta}}$$
$$\bar{y}_2 = \widehat{\gamma}_2 + \widehat{\beta}_2\bar{S}_2 + \widehat{\beta}_4\bar{Q}_2 + \bar{\mathbf{x}}_2\widehat{\boldsymbol{\beta}}$$

It follows that:

$$\bar{y}_1 - \bar{y}_2 = \widehat{\Delta}_1^* + \widehat{\beta}_2(\bar{S}_1 - \bar{S}_2) + \widehat{\beta}_4(\bar{Q}_1 - \bar{Q}_2) + (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\widehat{\boldsymbol{\beta}} \tag{3}$$

Because of the zero residuals property that holds in linear regressions, these hold identically. The left-hand-side of this familiar Oaxaca's decomposition is the raw differential, and is fixed. It decomposes into a conditional differential (which varies by specification) and a characteristics effect, made of three terms in this particular situation. The characteristics effect estimates the gap between the two differentials caused by the fact that g is (potentially) correlated with S, Q and $\mathbf{x}$. If there is no correlation, then $\bar{S}_1 = \bar{S}_2$, $\bar{Q}_1 = \bar{Q}_2$ and $\bar{\mathbf{x}}_1 = \bar{\mathbf{x}}_2$, and the two differentials coincide. This discussion should therefore clarify the confusion between interactions and correlations. Consider the selective school variable S. If $\beta_1 \neq \beta_2$, the effect of selective schooling on educational attainment varies between boys and girls, and this effect is seen in the conditional differential. This is not the same as $\bar{S}_1 \neq \bar{S}_2$, which says that selective schooling is correlated with gender. The latter effect is why the raw and conditional differentials might differ from each other.

We therefore decompose $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\widehat{\boldsymbol{\beta}}$ variable by variable to see which make any kind of contribution to the gap between the 2 differentials. If any variable does, it is because boys are different to girls, *and* because the variable has an impact on the attainment measure y.

Results

Tables 5, 6 and 7 summarise the regression results for the three dependent variables respectively.[6] Each table reports estimates, for YCS2 through to NPD2, for the "raw", "base" and "full" specifications. The corresponding tables of descriptive statistics are given in Tables 2, 3 and 4. The results are summarised in Figure 4.

From Figure 4, panel (a), we can see that the raw differential taken from published DfES data is always above the raw differential using YCS data except for the final two observations, where we are using NPD rather than YCS data, where the two differentials obviously coincide. In other words, the YCS has a downwards non-response bias, the reasons for which we need to document more clearly. This effect is the same when we use the regression samples, which are smaller because of missing values on many covariates.

In the same figure, we also plot the conditional gender gap ("full" specification). This conditional gap is about 1pp above the raw ("regn") gap, which implies being a girl is correlated with some covariates (see below). Panel (b) repeats for the Points Score measure, for which there are no corresponding published data. This

[6] Throughout we use robust standard errors, clustered on school s.

figure suggests that there is less difference between the raw and conditional gaps (0.5 of a GCSE grade) and that this difference disappears completely for the NPD.

What explains the gender gap (differential)?

The tables shown thus far summarise the regressions described above, which is Equation (1) and the differential Δ_1^*, computed from Equation (2). Table 8 reports these regressions in more detail, for various years (YCS3, YCS6, YCS10, NPD2).

The effect of single-sex schools and selective schooling on attainment at GCSE is very large and significant. For example, the marginal effect of selective schooling on Binary 5+ Passes in NPD2 is 0.690 log-points and is adds 12.4 points to the Point Score measure. Going to a single-sex school adds another 0.068 log-points to the first and 2.471 points to the second.

These variables do not affect the differential directly. The issue is whether these effects vary by gender and therefore contribute to the conditional gender differential. The effect of going to an all girls' school, over and above a single-sex school, is insignificant in almost all of the regressions, and when multiplied by $\bar{Q}_1$, makes no contribution whatsoever. The additional negative effect for girls going to selective schools, given by the estimates on gS, is not robustly estimated across year and attainment measure.[7]

Of all the possible interactions between g and every covariate we have access to, only one of these, the interaction gS, has any effect on the conditional differential (comparing the effect of g with the differential in the "base" specifications). Adding all the remaining covariates (individual, school, neighbourhood) changes the estimates very slightly, compare the "base" and "full" specifications. The differentials generally get bigger, by roughly one log-point in the Binary regressions and by just a tad in the Points Score regressions. (The effect of selective schooling falls a lot, ie pupils that go to selective schools are observably different from those that don't, but this doesn't affect the differential.)

Why do the raw and conditional differentials diverge? One can compare "raw" with "full", and use Oaxaca's decomposition, to examine exactly which variables are correlated with g and therefore affect the gap between the two (see Equation (3)).

[7] The strongest effect is in the Points Score regression for NPD2, where there is a negative effect at 2.4 points. This is a "ceiling" effect. Boys and girls who go to selective schools do as well as each other, because they tend to get the very top grades. Compared with boys and girls who don't go to selective schools, boys actually go further. With the tendency for grades to get better through time, this ceiling effect is more noticeable in later years.

Various variables are correlated with gender (whose effects are not reported in the tables). In the NPD, two explain why the conditional differential is lower, the third goes the other way. A higher proportion of girls go to selective schools, and as selective schools improve attainment, this effect is positive. Similarly, a higher proportion of boys than girls require special educational needs and obviously all special needs pupils have a lower attainment at GCSE. However, more girls than boys are eligible for free school meals, and so the effect goes the other way. This variable is intended to be a proxy for family background, which one might think should be uncorrelated with gender. In the YCS, being a girl is correlated with social housing, family background (e.g. father professional), and peer effects (the proportion 'poor' (YCS8 only) and the proportion with special educational needs (YCS10 only). All of these variables work in the same direction, and each contributes a little towards the gap between raw and conditional already noted.

The main conclusion of the paper thus far is that observables are not correlated with gender in a way that suggests that the conditional differential is much different from the raw differential (see Figure 4), and certainly does not explain why the differential has improved in favour of girls through the 1990s. There is no observable reason, given the data we have access to, that explains why the differential has moved in favour of girls as shown. If girls are not observably different from boys, girls must be behaving differently from boys at some point prior to taking their GCSEs. In the rest of the paper, we examine: (a) choice of 'secondary' school; (b) choice of subjects at GCSE; and (c) differences in exam performance between KS2 and KS4.

7.2 Choice of 'secondary' school

Above we have already reported that there are no significant interactions between gender and our vectors of school-level and neighbourhood covariates. One possible explanation for the gender gap is that girls' (by their parents or teachers) are being selected into unobservably better schools. To investigate this, we control for η_s using fixed-effects, labelled FE(s). To do this, we pool YCS6-11 and pool NPD1-2, as these are where the school identifier s is consistently defined within each pool. YCS2-5 are treated as four separate regressions.

The results are reported in Tables 9 to 11 and are plotted in Figure 4 for the Binary 5+ and Points Score measures. Roughly, the effect on the gender differential is roughly halved between 1991 and 1999, where it is possible that the trend has flattened out sightly. However, there is no effect in YCS2-5 or in the NPD. This is because having repeated observations on schools over time in YCS6-11 is important. It should be noted that the lines in Panel (a) are not strictly comparable because FE(s) is applied to the linear probability model, but the outcome is the same for the Points Score measure.

Clearly, between 1991 and 1999, $\widehat{\eta}_s$ and g_i are positively correlated. Because g_i is exogenous, we can regress $\widehat{\eta}_s$ on g_i (and time dummies) to obtain a differential of 0.033 (0.0029) log-points for Binary 5+, a differential of 0.031 (0.0032) log-points for Binary 8+, and a differential of 1.59 (0.129) GCSE points for Points Score. In other words, the average unobserved quality of schools that girls attend is 0.03 log-points better than those schools attended by boys.

We conclude that, in spite of there being lots of time-varying school-level covariates in the data, η_s is important. The most recent estimate (YCS10) of the gender gap, for the headline Binary 5+ measure, is 0.039 rather than 0.104. The quasi-privatisation of schooling, together with the introduction of school league tables in 1992, has possibly partly contributed to a spurious observed increase in the gender gap through the 1990s, simply because the better schools have selected better girls rather boys at the margin. The uncertainty in this conclusion- arises because we would ideally like to control for school fixed-effects in the late 1980s, but because the school identifier is not consistently recorded over this period, the results for this period are less credible.

7.3 Choice of subjects at GCSE

Table 12 dis-aggregates gender gaps by subject, by running cohort-specific regressions that have six subject dummies and six g–subject interactions. The first panel are raw differentials and the second panel conditional differentials. Table 13 repeats the exercise for 1991-1999, but also controlling for school-level fixed-effects as in the previous subsection. Figure 5 plots the first and fourth of these, that is compares raw differentials with conditional differentials that control for observables and unobservables, for the Points Score measure. The difference between the two is due to unobservables because it is clear that covariates have almost no effect.

In the raw data, girls outperformed boys in languages, English and vocational subjects throughout the sample period, overtook boys in humanities in the early 1990s, and have almost caught up with boys in maths and science. There are two noticeable features to these increasing raw gender gaps. The first is that there is a clear "one-off" GCE to GCSE effect that disadvantaged boys in languages, sciences and maths. In these subjects, the gender gap was falling until the change in exam regime in 1987. It is a "one-off" effect in that, since 1988, the gap has increased at the same rate in all subjects. (Note that there is a structural break in the series between 2001 and 2002.) This GCE to GCSE effect has been commented on by educationalists, and is almost certainly due to a change in the way pupils are examined in GCSE (ie more coursework, less exams). The fact the gap is betting bigger at the same rate means that the explanation has little to do the fact that girls are better at some subjects than others. It also suggests that the choice of subjects at GCSE (boys tend to prefer sciences, for example) does not provide an explanation.

Controlling for observables and unobservables does not alter any of the above, except that the gap is smaller by one-tenth of a GCSE grade in all subjects. This effect applies to all 6 subjects, except that there is a suggestion that the gap is actually closing in English and languages. This simply mimics the effect discussed in the previous subsection. In 2002, girls now ahead in three (English, languages, vocational), level in humanities, and behind in two (Maths and Science)

To conclude, the change from GCE to GCSE advantaged girls more than boys in some subjects, but since then the gap is getting wider in most subjects. There is no evidence that the gap is getting wider because girls are increasingly choosing subjects that advantage them the best.

7.4 Analysing Key Stage Exam performance using the NPD panel

The advantage of the NPD panel, compared with the pooled cross-section data, is that we observe educational attainment more than once, and therefore we can control for and estimate α_i as well as η_s. Specifically, for each pupil i, we observe exam results (so-called SAT scores) y at Key Stage 2, Key Stage 3, Key Stage 4 (ie GCSE) ($t = 2, 3, 4$) for 3 subjects: English, Maths, Science ($k = 1, 2, 3$) Thus, for each subject, we estimate

$$y_{it} = \gamma_2 + \gamma_3 k_t^3 + \gamma_4 k_t^4 + \beta_2 g_i + \beta_3 k_t^3 g_i + \beta_4 k_t^4 g_i + \mathbf{x}\boldsymbol{\beta} + \alpha_i + \eta_s + u_{it}$$

where k^2, k^3, k^4 are dummies for Key Stage.

Without α_i, η_s and $\mathbf{x}$,

$$\beta_2 \equiv d_2 = \mathrm{E}(y \mid k = 2, g = 1) - \mathrm{E}(y \mid k = 2, g = 0)$$

ie the gender gap at Key Stage 2, and

$$\beta_3 = d_3 - d_2 \qquad \beta_4 = d_4 - d_2$$

show how the gender gap compares with Key Stage 2. In other words, we can analyse how these gaps evolve over the pupils' secondary education, and whether they vary by subject. The Key Stage scores ('levels') are meant to be comparable between Key Stages 3 and 4, whereas pupils are meant to progress by one level between Key Stages 2 and 3. The ensure comparability between all three Key Stages, we therefore add unity to Key Stage 2.

To see how exam results evolve over the three Key Stage exams, the following table cross-tabulates Maths scores between Key Stage 2 (y_{i2}) and Key Stage 3 (y_{i3}):

	y_{i2}					
y_{i3}	2	3	4	5	6	Total
2	5,660	1,102	408	32	0	7,202
3	17,479	19,476	1,603	82	2	38,643
4	6,133	69,969	20,740	490	3	97,335
5	318	39,645	80,342	3,285	8	123,598
6	66	5,128	93,637	21,460	22	120,313
7	43	203	28,725	51,829	118	80,918
8	4	19	586	12,133	345	13,087
9	0	0	4	92	54	150
Total	29,703	135,542	226,045	89,403	552	481,246

Most pupils stay the same, or move up one Key Stage level. For example, the modal outcome at Key Stage 2 is Level 4, and by far the biggest majority achieve Level 5 or 6 at Key Stage 3.

	y_{i3}								
y_{i4}	2	3	4	5	6	7	8	9	Total
0	3,400	6,845	4,959	2,735	1,032	213	4	0	19,188
1	1,651	11,457	5,558	559	43	2	0	0	19,270
2	680	15,844	25,621	2,786	56	6	0	0	44,993
3	261	4,310	43,558	26,738	1,770	38	0	0	76,675
4	161	346	17,181	52,552	13,749	197	1	0	84,187
5	109	26	2,238	35,169	59,568	8,927	17	0	106,054
6	27	11	187	5,204	42,932	37,689	545	4	86,599
7	6	10	12	54	2,957	29,577	5,214	13	37,843
8	7	9	12	4	101	5,140	7,404	134	12,811
Total	6,302	38,858	99,326	125,801	122,208	81,789	13,185	151	487,620

The same metric used for Key Stages 3 and 4. It is noticeable that a lot of pupils actually go backwards between the two Key Stages.

By construction, the NPD does not record which 'primary' school was attended at Key Stage 2. Thus we assume that $s_2 = s_3$. However, some pupils change school between Key Stages 3 and 4, so pupils can either have one or two *spells* l in each school:

keystage	2	3	4
one spell	1	1	1
two spells	1	1	2

In what follows we use NPD1 only, and construct a balanced panel in that we observe y at $k = 2, 3, 4$ for all three subjects. The following table summarises the information:

Sample	
Schools s	3,090
Obs it	1,260,996
Pupils i	422,524
Spells l	430,326
Pupils with 2 spells	7,802
Obs/pupil	3
Pupils/school	137

We estimate the following four econometric models:

1. Pooled OLS (neither α_i or η_s are identified).

2. FE(i). This model controls for and estimates α_i, but the estimates on g, $\mathbf{x}_i$ are not identified. For the effects of $\mathbf{x}_s$ to be identified, we need either $\mathbf{x}_{s(it)t}$ need to vary over time or individuals change school.

3. FE(s). This model controls for and estimates η_s. The estimates on $\mathbf{x}_{s(i)}$ are not identified, but for the estimates on $\mathbf{x}_{s(it)t}$ to be identified, the school variables need to vary over time.

4. FE(l) This model controls for α_i and η_s, but we cannot estimate them, only their sum. This is because we define the heterogeneity at the spell level and sweep out by creating spell-level mean deviations. This is an important issue in the econometrics of matched employee-employer data: if we want to estimate corr(α_i, η_s), then one should use double fixed-effects methods. However, we have too many schools, and we also believe that correlation is biased downwards (Andrews, Schank & Upward 2005*b*, Andrews, Schank & Upward 2005*a*). So the correlations we report below are from FE(i) and FE(s) separately.

Results

The results are shown in Table 14. Looking at the Pooled OLS results first, it can be seen that the biggest gap at GCSE is in English at 0.678 (=0.214+0.464) of a KS level, compared with Maths at 0.074 (=0.182-0.108) of a KS level (0.182-0.108), and Science at 0.092 (=0.163-0.071) of a KS level. This, of course, confirms what is well-known elsewhere, as discussed earlier in this paper. What is less well-known is how these gaps differ at ages 11 and 14. Although girls are better in English at Key Stage 2 (0.214), behind in Maths (-0.108) and Science (-0.071), the gaps between the three subjects are a lot smaller. In other words, girls improve between Key Stage 3 and Key Stage 4 in all 3 subjects, but only in English between Key Stage 2 and Key Stage 3. Boys only improve slightly in Science between Key Stage 2 and Key Stage 3.

The correlations between the estimates of α_i [from FE(i)], η_s [from FE(s)] and g are given in the following table:

	α_1	α_2	α_3	α	η	g
α_1 (eng)	1.0000					
α_2 (mat)	0.7948	1.0000				
α_3 (sci)	0.8190	0.8791	1.0000			
α	0.9181	0.9518	0.9555	1.0000		
η	0.3714	0.3745	0.4326	0.4167	1.0000	
g	0.0902	-0.0734	-0.0291	-0.0093	0.0636	1.0000

Three things stand out. First, there is a strong correlation of 0.42 between $\widehat{\alpha}_i$ and $\widehat{\eta}_s$: unobservably good pupils go to unobservably good schools. When $\widehat{\alpha}_i$ is decomposed by subject, this correlation is weakest for English and Maths, but strongest in Science. Finally, there is a weak correlation between g and η, which means that girls do go to unobservably better schools, but this result is less convincing that reported earlier in the YCS. This is probably because we need to analyse more sweeps of the NPD.

8 Conclusions

Our main findings are:

1. In the raw data the gender gap widened considerably following the introduction of the GCSE exams in 1987. It continued to widen quite rapidly until the early

1990s and eventually stabilised at the end of the 1990s. By 2000, for example, there was a ten percentage point gap between girls and boys in the proportion gaining 5 or more A*-C grades in the GCSE exams.

2. We test a large number of hypotheses that have been suggested as 'causes' of the gender gap. This involves estimating econometric models where we control for a large number of observable factors. These factors can be grouped into personal (e.g. ethnicity), family (e.g. socio-economic background), school (e.g. selective school, single sex) and environmental (e.g. local unemployment rate). Controlling for such factors fails to explain the gender gap.

   (a) For instance, selective schools have a very large effect on educational outcomes, whereas single sex schools have a smaller effect, but neither of these observable school-level effects can explain the gender gap.

   (b) Furthermore, there are no observable differences between girls and boys (e.g. family background, poverty), and hence these variables do not explain why there is a gender gap, or why it has risen.

3. In view of these findings, we argue that girls must behave differently to boys prior to the GCSE stage. Consequently, we explore (a) the effect of secondary school choice on the gender gap, (b) subject-level differences in the gap and (c) differences in exam performance between Key Stages 2 and 4. In this part of our research we are also able to control, statistically, for factors that are unobserved in our data, such as the school's ethos. Controlling for unobserved factors proves to be important in reducing the size of the gap but does not explain its upward trend. Thus, between 1991 and 2001 one-half of the gender gap can be explained by unobserved differences between schools. Using the data for 2002, for example, the gap falls from 10 percentage points to 4 percentage points when we control for unobserved school effects. We therefore conclude that unobservable differences between schools, which could include variables such as pupil behaviour, tiering and streaming, could well be important explanations of the gender gap even though we have no direct evidence of these effects.

4. With respect to the gender gap at subject level, the raw data show that girls substantially outperform boys in languages, English and vocational subjects, such as Business Studies and to a lesser extent in humanities. Girls have also caught up with boys in Science and Maths after being well behind boys in these subjects at the start of the GCSE exams in 1987. When we control for

observable and unobservable (school-level) differences between individuals, the gender gap is reduced by one-tenth of a GCSE grade, so we find that girls are still way ahead of boys in English, languages and vocational subjects, but are slightly behind boys in Maths and Science.

5. Our analysis of the changes in test scores between different stages of the educational process shows that by the time that pupils take their GCSE exams, girls are ahead of boys by nearly two thirds of a grade in English, but are only slightly ahead in Maths and Science. However, girls are already well ahead in English by Key Stage 2, but behind in Maths and Science, which means that girls improve relative to boys between Key Stages 3 and 4 in all subjects, but only in English between Key Stages 2 and 3.

Our overall conclusion is that the quasi-privatisation of schooling, together with the introduction of school league tables in 1992, has possibly partly contributed to a spurious observed increase in the gender gap through the 1990s, simply because the better schools have selected better girls rather boys at the margin.

Tables

Table 1: The panel structure of the NPD[a]

cohort 1	cohort 2	KS	year	age	data
1996/7	1997/8	2	6	11	y_{i2}, $\mathbf{x}_i$
1999/0	2000/1	3	9	14	y_{i3}, $\mathbf{x}_i$, $\mathbf{x}_{s(i3)3}$, $\mathbf{x}_{r(i3)3}$
2001/2	2002/3	4	11	16	y_{i4}, $\mathbf{x}_i$, $\mathbf{x}_{s(i4)4}$, $\mathbf{x}_{r(i4)4}$

[a] Key Stage (KS) defines the time dimension $t = 2, 3, 4$.

Table 2: Descriptive statistics, Binary 5+ Passes (5+ $A^{*}-C$ GCSEs)

cohort	year	N[a]	$\bar{g}$[b]	$\bar{y}$[c]	$\bar{y}_g$[d]	$\bar{y}_b$[e]	raw[f]	prraw[g]	$\bar{z}_g$[h]	$\bar{z}_b$[i]
				DfES (population)[j]						
	1979	737.3	0.489	0.236	0.239	0.234	0.005	0.021		
	1985	736.2	0.491	0.268	0.274	0.263	0.011	0.042		
	1986	718.2	0.491	0.267	0.272	0.262	0.010	0.038		
	1988	656.0	0.488	0.299	0.317	0.282	0.035	0.124		
	1990	582.1	0.487	0.345	0.384	0.308	0.076	0.247		
	1991	555.2	0.500	0.368	0.403	0.333	0.070	0.210		
	1993	512.1	0.489	0.412	0.458	0.368	0.090	0.245		
	1995	528.3	0.495	0.435	0.481	0.390	0.091	0.233		
	1997	537.6	0.484	0.451	0.500	0.405	0.095	0.235		
	1999	581.0	0.481	0.479	0.534	0.428	0.106	0.248		
	2001	603.4	0.491	0.500	0.554	0.448	0.106	0.237		
	2002	510.4	0.491	0.516	0.570	0.464	0.106	0.228		
	2003	534.1	0.485	0.529	0.582	0.479	0.103	0.215		
				YCS/NPD (full sample)						
YCS2	1985	12807	0.533	0.265	0.269	0.262	0.007	0.027	7.13	6.83
YCS3	1986	13537	0.508	0.259	0.256	0.261	-0.005	-0.019	7.03	6.76
YCS4	1988	12085	0.523	0.314	0.321	0.307	0.014	0.046	6.59	6.31
YCS5	1990	11712	0.530	0.398	0.421	0.372	0.049	0.132	7.09	6.80
YCS6	1991	18737	0.542	0.421	0.440	0.398	0.042	0.106	7.69	7.38
YCS7	1993	15442	0.547	0.482	0.509	0.449	0.060	0.134	8.21	7.86
YCS8	1995	13136	0.546	0.549	0.575	0.517	0.058	0.112	8.82	8.60
YCS9	1997	12331	0.535	0.564	0.594	0.529	0.065	0.123	8.75	8.57
YCS10	1999	11467	0.541	0.617	0.652	0.576	0.076	0.132	8.96	8.68
YCS11	2001	13997	0.556	0.629	0.669	0.578	0.091	0.157	8.46	8.31
NPD1	2002	510360	0.495	0.498	0.551	0.446	0.105	0.235	8.66	8.33
NPD2	2003	534132	0.494	0.491	0.542	0.440	0.102	0.232	8.56	8.18
				YCS/NPD (regression sample)[k]						
YCS2	1985	6625	0.538	0.307	0.304	0.311	-0.007	-0.023	7.47	7.21
YCS3	1986	5557	0.535	0.345	0.338	0.354	-0.016	-0.045	7.53	7.35
YCS4	1988	4188	0.566	0.451	0.449	0.454	-0.005	-0.011	7.25	7.11
YCS5	1990	10586	0.533	0.398	0.419	0.374	0.045	0.120	7.09	6.82
YCS6	1991	16179	0.537	0.450	0.475	0.421	0.054	0.128	7.89	7.55
YCS7	1993	11874	0.548	0.501	0.526	0.470	0.056	0.119	8.31	7.96
YCS8	1995	9995	0.545	0.571	0.601	0.534	0.067	0.125	8.94	8.75
YCS9	1997	6468	0.537	0.582	0.612	0.546	0.066	0.121	8.82	8.63
YCS10	1999	7926	0.546	0.679	0.713	0.638	0.075	0.118	9.22	8.96
YCS11	2001	8225	0.556	0.685	0.729	0.631	0.098	0.155	8.57	8.42
NPD1	2002	475370	0.496	0.498	0.549	0.448	0.101	0.225	8.73	8.45
NPD2	2003	491564	0.496	0.491	0.541	0.441	0.100	0.227	8.62	8.30

[a] number of observations (population figures in thousands)
[b] proportion of girls
[c] proportion of passes
[d] proportion of passes for girls
[e] proportion of passes for boys
[f] absolute gender gap, $\bar{y}_g - \bar{y}_b$
[g] relative gender gap, $\bar{y}_g / \bar{y}_b - 1$
[h] number of exams for girls
[i] number of exams for boys
[j] Published data from 'Statistics of Education: Public Exams, GCSE and GCE in England' (various years), 'DES Statistical Bulletin 1/91, School Exam Survey 1988-89', 'Regional Trends' (1999,2000,2001) and the DfES website http://www.dfes.gov.uk/trends
[k] observations dropped due to missing covariates

Table 3: Descriptive statistics, Binary 8+ Passes (8+ $A^{*} - C$ GCSEs)

cohort	year	N[a]	$\bar{g}$[b]	$\bar{y}$[c]	$\bar{y}_g$[d]	$\bar{y}_b$[e]	raw[f]	prraw[g]	$\bar{z}_g$[h]	$\bar{z}_b$[i]
				YCS/NPD (full sample)						
YCS2	1985	12807	0.533	0.105	0.111	0.098	0.013	0.133	7.13	6.83
YCS3	1986	13537	0.508	0.108	0.108	0.108	0.000	0.000	7.03	6.76
YCS4	1988	12085	0.523	0.113	0.117	0.108	0.009	0.083	6.59	6.31
YCS5	1990	11712	0.530	0.180	0.191	0.168	0.023	0.137	7.09	6.80
YCS6	1991	18737	0.542	0.226	0.249	0.199	0.050	0.251	7.69	7.38
YCS7	1993	15442	0.547	0.286	0.314	0.251	0.063	0.251	8.21	7.86
YCS8	1995	13136	0.546	0.367	0.397	0.332	0.065	0.196	8.82	8.60
YCS9	1997	12331	0.535	0.369	0.409	0.323	0.086	0.266	8.75	8.57
YCS10	1999	11467	0.541	0.436	0.483	0.380	0.103	0.271	8.96	8.68
YCS11	2001	13997	0.556	0.385	0.431	0.327	0.104	0.318	8.46	8.31
NPD1	2002	510360	0.495	0.335	0.389	0.282	0.107	0.379	8.66	8.33
NPD2	2003	534132	0.494	0.323	0.377	0.270	0.107	0.396	8.56	8.18
				YCS/NPD (regression sample)[j]						
YCS2	1985	6625	0.538	0.122	0.123	0.119	0.004	0.034	7.47	7.21
YCS3	1986	5557	0.535	0.153	0.149	0.158	-0.009	-0.057	7.53	7.35
YCS4	1988	4188	0.566	0.180	0.183	0.176	0.007	0.040	7.25	7.11
YCS5	1990	10586	0.533	0.177	0.187	0.166	0.021	0.127	7.09	6.82
YCS6	1991	16179	0.537	0.247	0.275	0.215	0.060	0.279	7.89	7.55
YCS7	1993	11874	0.548	0.302	0.331	0.267	0.064	0.240	8.31	7.96
YCS8	1995	9995	0.545	0.385	0.414	0.350	0.064	0.183	8.94	8.75
YCS9	1997	6468	0.537	0.385	0.428	0.334	0.094	0.281	8.82	8.63
YCS10	1999	7926	0.546	0.492	0.544	0.431	0.113	0.262	9.22	8.96
YCS11	2001	8225	0.556	0.432	0.484	0.368	0.116	0.315	8.57	8.42
NPD1	2002	475370	0.496	0.331	0.384	0.280	0.104	0.371	8.72	8.45
NPD2	2003	491564	0.496	0.319	0.372	0.266	0.106	0.398	8.62	8.30

[a] number of observations
[b] proportion of girls
[c] proportion of passes
[d] proportion of passes for girls
[e] proportion of passes for boys
[f] absolute gender gap, $\bar{y}_g - \bar{y}_b$
[g] relative gender gap, $\bar{y}_g/\bar{y}_b - 1$
[h] number of exams for girls
[i] number of exams for boys
[j] observations dropped due to missing covariates

Table 4: Descriptive statistics, Points Score

cohort	year	N[a]	$\bar{g}$[b]	$\bar{y}$[c]	$\bar{y}_g$[d]	$\bar{y}_b$[e]	raw[f]	prraw[g]	$\bar{z}_g$[h]	$\bar{z}_b$[i]
					YCS/NPD (full sample)					
YCS2	1985	12807	0.533	28.530	29.236	27.724	1.512	0.053	7.13	6.83
YCS3	1986	13537	0.508	28.199	28.663	27.719	0.944	0.033	7.03	6.76
YCS4	1988	12085	0.523	26.552	27.262	25.773	1.489	0.056	6.59	6.31
YCS5	1990	11712	0.530	30.496	31.523	29.340	2.183	0.072	7.09	6.80
YCS6	1991	18737	0.542	32.712	33.807	31.414	2.393	0.073	7.69	7.38
YCS7	1993	15442	0.547	35.826	37.269	34.083	3.186	0.089	8.21	7.86
YCS8	1995	13136	0.546	40.496	41.850	38.869	2.981	0.074	8.82	8.60
YCS9	1997	12331	0.535	40.655	42.062	39.036	3.026	0.075	8.75	8.57
YCS10	1999	11467	0.541	43.330	45.088	41.254	3.834	0.089	8.96	8.68
YCS11	2001	13997	0.556	41.814	43.200	40.075	3.125	0.075	8.46	8.31
NPD1	2002	510360	0.495	38.311	40.722	35.950	4.772	0.125	8.66	8.33
NPD2	2003	534132	0.494	37.566	40.042	35.151	4.891	0.130	8.56	8.18
					YCS/NPD (regression sample)[j]					
YCS2	1985	6625	0.538	30.816	31.260	30.300	0.960	0.031	7.47	7.21
YCS3	1986	5557	0.535	32.027	32.166	31.866	0.300	0.009	7.53	7.35
YCS4	1988	4188	0.566	32.427	32.555	32.261	0.294	0.009	7.25	7.11
YCS5	1990	10586	0.533	30.533	31.447	29.488	1.959	0.064	7.09	6.82
YCS6	1991	16179	0.537	34.104	35.363	32.642	2.721	0.080	7.89	7.55
YCS7	1993	11874	0.548	36.670	38.127	34.906	3.221	0.088	8.31	7.96
YCS8	1995	9995	0.545	41.580	42.936	39.954	2.982	0.072	8.94	8.75
YCS9	1997	6468	0.537	41.436	42.783	39.876	2.907	0.070	8.82	8.63
YCS10	1999	7926	0.546	46.093	47.859	43.969	3.890	0.085	9.22	8.96
YCS11	2001	8225	0.556	43.706	45.090	41.974	3.116	0.072	8.57	8.42
NPD1	2002	475370	0.496	38.516	40.795	36.277	4.518	0.117	8.72	8.45
NPD2	2003	491564	0.496	37.758	40.125	35.430	4.695	0.124	8.62	8.30

[a] number of observations
[b] proportion of girls
[c] average points score
[d] average points score for girls
[e] average points score for boys
[f] absolute gender gap, $\bar{y}_g - \bar{y}_b$
[g] relative gender gap, $log(\bar{y}_g/\bar{y}_b)$
[h] number of exams for girls
[i] number of exams for boys
[j] observations dropped due to missing covariates

Table 5: Summary of regression results, Binary 5+ Passes (5+ $A^* - C$ GCSEs)[*]

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	-0.007 (0.012)	-0.021 (0.012)	-0.015 (0.012)
YCS3	1986	5557	-0.015 (0.015)	-0.009 (0.015)	-0.002 (0.014)
YCS4	1988	4188	-0.005 (0.018)	0.000 (0.017)	0.009 (0.018)
YCS5	1990	10585	0.045 (0.013)	0.048 (0.011)	0.062 (0.011)
YCS6	1991	16179	0.053 (0.009)	0.052 (0.010)	0.069 (0.011)
YCS7	1993	11874	0.056 (0.010)	0.039 (0.011)	0.057 (0.012)
YCS8	1995	9995	0.066 (0.011)	0.070 (0.011)	0.095 (0.012)
YCS9	1997	6468	0.066 (0.013)	0.064 (0.014)	0.081 (0.014)
YCS10	1999	7926	0.075 (0.011)	0.092 (0.018)	0.104 (0.017)
YCS11	2001	8225	0.098 (0.011)	0.100 (0.012)	0.118 (0.012)
NPD1	2002	482967	0.103 (0.004)	0.105 (0.003)	0.117 (0.003)
NPD2	2003	503436	0.100 (0.004)	0.101 (0.003)	0.114 (0.003)

[*] cohort-specific regressions using dlogit2

[a] girl dummy

[b] raw plus month of birth dummies (except YCS8-10, for which only year of birth is available), girls-only school, boys-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available)

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

Table 6: Summary of regression results, Binary 8+ Passes (8+ $A^* - C$ GCSES)*

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	0.004 (0.008)	0.000 (0.008)	0.003 (0.007)
YCS3	1986	5557	-0.009 (0.012)	-0.007 (0.010)	-0.003 (0.009)
YCS4	1988	4188	0.007 (0.013)	0.013 (0.012)	0.017 (0.011)
YCS5	1990	10585	0.020 (0.011)	0.025 (0.008)	0.027 (0.006)
YCS6	1991	16179	0.060 (0.008)	0.060 (0.007)	0.066 (0.007)
YCS7	1993	11874	0.064 (0.010)	0.059 (0.010)	0.070 (0.009)
YCS8	1995	9995	0.064 (0.011)	0.069 (0.011)	0.090 (0.011)
YCS9	1997	6468	0.095 (0.013)	0.103 (0.013)	0.125 (0.014)
YCS10	1999	7926	0.113 (0.013)	0.121 (0.012)	0.147 (0.013)
YCS11	2001	8225	0.117 (0.012)	0.121 (0.012)	0.147 (0.013)
NPD1	2002	482967	0.106 (0.004)	0.110 (0.003)	0.117 (0.002)
NPD2	2003	503436	0.107 (0.004)	0.110 (0.003)	0.118 (0.002)

* cohort-specific regressions using dlogit2

[a] girl dummy

[b] raw plus month of birth dummies (except YCS8-10, for which only year of birth is available), girls-only school, boys-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available)

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

Table 7: Summary of regression results, Points Score*

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	0.961 (0.358)	0.663 (0.344)	0.866 (0.314)
YCS3	1986	5557	0.299 (0.464)	0.459 (0.403)	0.738 (0.371)
YCS4	1988	4188	0.294 (0.529)	0.547 (0.485)	0.860 (0.419)
YCS5	1990	10585	1.959 (0.466)	2.068 (0.353)	2.367 (0.312)
YCS6	1991	16179	2.722 (0.315)	2.605 (0.285)	2.887 (0.245)
YCS7	1993	11874	3.221 (0.369)	2.943 (0.344)	3.229 (0.282)
YCS8	1995	9995	2.982 (0.378)	3.094 (0.351)	3.591 (0.308)
YCS9	1997	6468	2.908 (0.465)	2.960 (0.432)	3.250 (0.373)
YCS10	1999	7926	3.890 (0.430)	3.804 (0.379)	4.054 (0.344)
YCS11	2001	8225	3.115 (0.355)	3.119 (0.314)	3.481 (0.289)
NPD1	2002	482967	4.540 (0.152)	4.406 (0.116)	4.501 (0.085)
NPD2	2003	503436	4.651 (0.158)	4.512 (0.118)	4.621 (0.085)

* cohort-specific regressions using OLS

[a] girl dummy

[b] raw plus month of birth dummies (except YCS8-10, for which only year of birth is available), girls-only school, boys-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available)

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

Table 8: Summary of regression results*

	YCS3 (1986)		YCS6 (1991)		YCS10 (1999)		NPD2 (2003)	
	base[a]	full[b]	base	full	base	full	base	full
					Binary 5+ Passes (5+ $A^* - C$ GCSEs)			
gender g	-0.014 (0.015)	-0.005 (0.014)	0.056 (0.009)	0.074 (0.009)	0.064 (0.010)	0.078 (0.011)	0.100 (0.002)	0.111 (0.002)
single sex school Q			0.051 (0.031)	0.103 (0.034)	0.007 (0.033)	0.035 (0.032)	0.055 (0.018)	0.057 (0.013)
all girls' school G			-0.030 (0.040)	-0.037 (0.040)	0.060 (0.042)	0.062 (0.041)	0.005 (0.023)	0.028 (0.015)
selective school S	0.566 (0.065)	0.518 (0.062)	0.689 (0.130)	0.523 (0.130)	0.685 (0.109)	0.509 (0.105)	0.920 (0.031)	0.724 (0.029)
interaction gS	0.117 (0.102)	0.099 (0.100)	-0.006 (0.142)	-0.007 (0.136)	0.263 (0.226)	0.239 (0.212)	0.015 (0.040)	0.001 (0.038)
differential	-0.009 (0.015)	-0.002 (0.014)	0.052 (0.010)	0.069 (0.011)	0.092 (0.018)	0.104 (0.017)	0.101 (0.003)	0.114 (0.003)
raw differential	-0.015 (0.015)		0.053 (0.009)		0.075 (0.011		0.100 (0.004)	
					Binary 8+ Passes (8+ $A^* - C$ GCSEs)			
gender g	-0.008 (0.011)	-0.004 (0.009)	0.066 (0.007)	0.072 (0.007)	0.113 (0.012)	0.140 (0.013)	0.112 (0.002)	0.117 (0.002)
single sex school Q			0.059 (0.024)	0.092 (0.023)	0.028 (0.039)	0.060 (0.039)	0.061 (0.018)	0.062 (0.012)
all girls' school G			-0.078 (0.032)	-0.075 (0.028)	0.040 (0.050)	0.043 (0.049)	-0.013 (0.022)	0.011 (0.014)
selective school S	0.269 (0.027)	0.216 (0.025)	0.373 (0.040)	0.223 (0.036)	0.593 (0.054)	0.386 (0.058)	0.614 (0.023)	0.423 (0.020)
interaction gS	0.033 (0.035)	0.023 (0.032)	0.073 (0.050)	0.071 (0.043)	0.031 (0.078)	0.019 (0.075)	0.009 (0.029)	-0.005 (0.024)
differential	-0.007 (0.010)	-0.003 (0.009)	0.060 (0.007)	0.066 (0.007)	0.121 (0.012)	0.147 (0.013)	0.110 (0.003)	0.118 (0.002)
raw differential	-0.009 (0.012)		0.060 (0.008)		0.113 (0.013		0.107 (0.004)	
					Points Score			
gender g	0.435 (0.411)	0.734 (0.378)	2.814 (0.282)	3.065 (0.250)	3.863 (0.400)	4.105 (0.373)	4.689 (0.067)	4.687 (0.060)
single sex school Q			2.800 (1.084)	4.122 (0.930)	1.832 (1.187)	2.290 (1.008)	2.544 (0.745)	2.114 (0.498)
all girls' school G			-1.695 (1.376)	-1.483 (1.077)	1.245 (1.485)	1.316 (1.197)	-0.176 (0.928)	0.620 (0.581)
selective school S	16.855 (1.733)	14.494 (1.728)	17.286 (1.545)	9.138 (1.376)	16.274 (1.256)	8.827 (1.175)	22.829 (0.754)	14.222 (0.654)
interaction gS	0.625 (2.103)	0.109 (2.097)	-0.156 (1.823)	-0.004 (1.413)	-3.315 (1.614)	-3.341 (1.346)	-3.868 (0.949)	-3.649 (0.736)
differential	0.459 (0.403)	0.738 (0.371)	2.605 (0.285)	2.887 (0.245)	3.804 (0.379)	4.054 (0.344)	4.512 (0.118)	4.621 (0.085)
raw differential	0.299 (0.464)		2.722 (0.315)		3.890 (0.430		4.651 (0.158)	

* cohort-specific regressions using dlogit2 (Binary 5+ Passes and Binary 8+ Passes) and OLS (Points Score)

[a] girl dummy plus month of birth dummies (except YCS10, for which only year of birth is available), girls-only school, boys-only school, selective school, girl in selective school (except YCS3, for which information on single-sex schools is not available)

[b] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS3) and regional variables (except YCS3, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

Table 9: Summary of fixed effects linear probability model, Binary 5+ Passes (5+ $A^{*} - C$ GCSEs)*

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	-0.017 (0.014)	-0.020 (0.014)	-0.013 (0.014)
YCS3	1986	5557	-0.013 (0.014)	-0.010 (0.014)	-0.004 (0.014)
YCS4	1988	4188	0.008 (0.017)	0.008 (0.017)	0.011 (0.016)
YCS5	1990	10586	0.054 (0.010)	0.055 (0.010)	0.059 (0.010)
YCS6	1991	16179	0.055 (0.008)	0.041 (0.010)	0.044 (0.010)
YCS7	1993	11874	0.053 (0.009)	0.040 (0.012)	0.041 (0.011)
YCS8	1995	9995	0.074 (0.010)	0.059 (0.012)	0.064 (0.012)
YCS9	1997	6468	0.067 (0.012)	0.051 (0.015)	0.054 (0.014)
YCS10	1999	7926	0.081 (0.011)	0.062 (0.013)	0.064 (0.013)
YCS11	2001	8225	0.097 (0.010)	0.084 (0.012)	0.094 (0.012)
NPD1	2002	482967	0.100 (0.001)	0.097 (0.001)	0.099 (0.001)
NPD2	2003	503436	0.098 (0.001)	0.095 (0.001)	0.096 (0.001)

* YCS2, YCS3, YCS4, YCS5, YCS6-11, NPD1-2 pooled regressions

[a] girl dummy

[b] raw plus girls-only school, boys-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available); selective school is dropped for YCS2, YCS3, YCS4 and YCS5, girls-only school, boys-only school and selective school are dropped for NPD1-2

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing); secondary modern and all 3 regional variables are dropped for YCS2, YCS3, YCS4 and YCS5, expenditure is dropped for NPD1-2

Table 10: Summary of fixed effects linear probability model, Binary 8+ Passes (8+ $A^* - C$ GCSEs)*

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	0.001 (0.010)	-0.005 (0.011)	-0.001 (0.011)
YCS3	1986	5557	-0.006 (0.010)	-0.001 (0.011)	0.004 (0.011)
YCS4	1988	4188	0.011 (0.013)	0.011 (0.013)	0.012 (0.012)
YCS5	1990	10586	0.026 (0.007)	0.031 (0.009)	0.033 (0.008)
YCS6	1991	16179	0.063 (0.007)	0.032 (0.014)	0.034 (0.013)
YCS7	1993	11874	0.062 (0.008)	0.025 (0.017)	0.026 (0.016)
YCS8	1995	9995	0.072 (0.010)	0.039 (0.016)	0.042 (0.016)
YCS9	1997	6468	0.096 (0.012)	0.057 (0.019)	0.058 (0.018)
YCS10	1999	7926	0.115 (0.011)	0.076 (0.019)	0.078 (0.018)
YCS11	2001	8225	0.114 (0.011)	0.084 (0.016)	0.091 (0.015)
NPD1	2002	482967	0.103 (0.001)	0.102 (0.001)	0.103 (0.001)
NPD2	2003	503436	0.104 (0.001)	0.103 (0.001)	0.104 (0.001)

* YCS2, YCS3, YCS4, YCS5, YCS6-11, NPD1-2 pooled regressions

[a] girl dummy

[b] raw plus single-sex school, girls-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available); selective school is dropped for YCS2, YCS3, YCS4 and YCS5, single-sex school, girls-only school and selective school are dropped for NPD1-2

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing); secondary modern and all 3 regional variables are dropped for YCS2, YCS3, YCS4 and YCS5, expenditure is dropped for NPD1-2

Table 11: Summary of fixed effects linear regression results, Points Score*

cohort	year	N	raw[a]	base[b]	full[c]
YCS2	1985	6625	0.690 (0.403)	0.590 (0.406)	0.826 (0.386)
YCS3	1986	5557	0.523 (0.410)	0.596 (0.413)	0.819 (0.401)
YCS4	1988	4188	0.673 (0.496)	0.664 (0.500)	0.824 (0.458)
YCS5	1990	10586	2.306 (0.311)	2.295 (0.321)	2.487 (0.297)
YCS6	1991	16179	2.728 (0.257)	2.095 (0.318)	2.201 (0.299)
YCS7	1993	11874	3.061 (0.299)	2.458 (0.369)	2.519 (0.346)
YCS8	1995	9995	3.244 (0.327)	2.558 (0.389)	2.760 (0.367)
YCS9	1997	6468	2.956 (0.392)	2.222 (0.455)	2.348 (0.430)
YCS10	1999	7926	4.151 (0.356)	3.364 (0.427)	3.432 (0.408)
YCS11	2001	8225	2.997 (0.304)	2.445 (0.356)	2.825 (0.340)
NPD1	2002	482967	4.455 (0.046)	4.399 (0.046)	4.458 (0.045)
NPD2	2003	503436	4.605 (0.046)	4.542 (0.046)	4.614 (0.045)

* YCS2, YCS3, YCS4, YCS5, YCS6-11, NPD1-2 pooled regressions

[a] girl dummy

[b] raw plus girls-only school, boys-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available); selective school is dropped for YCS2, YCS3, YCS4 and YCS5, girls-only school, boys-only school and selective school are dropped for NPD1-2

[c] base plus school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing); secondary modern and all 3 regional variables are dropped for YCS2, YCS3, YCS4 and YCS5, expenditure is dropped for NPD1-2

Table 12: Summary of subject-specific regression results, Points Score*

cohort	year	N	English[c]	maths[d]	science[e]	humanities[f]	languages[g]	other[h]
					raw[a]			
YCS2	1985	24589	0.286 (0.040)	-0.380 (0.046)	-0.123 (0.073)	-0.131 (0.053)	0.360 (0.070)	
YCS3	1986	20669	0.297 (0.048)	-0.403 (0.057)	-0.023 (0.091)	-0.163 (0.062)	0.339 (0.081)	
YCS4	1988	19108	0.364 (0.054)	-0.561 (0.067)	-0.423 (0.070)	-0.050 (0.073)	0.246 (0.090)	0.242 (0.084)
YCS5	1990	50711	0.497 (0.045)	-0.344 (0.052)	-0.188 (0.050)	0.064 (0.061)	0.468 (0.081)	0.234 (0.055)
YCS6	1991	80142	0.512 (0.029)	-0.244 (0.035)	-0.193 (0.035)	0.157 (0.040)	0.615 (0.045)	0.325 (0.041)
YCS7	1993	61196	0.493 (0.032)	-0.166 (0.040)	-0.161 (0.038)	0.194 (0.045)	0.613 (0.049)	0.366 (0.043)
YCS8	1995	54665	0.524 (0.034)	-0.179 (0.040)	-0.195 (0.035)	0.168 (0.045)	0.569 (0.047)	0.419 (0.042)
YCS9	1997	34069	0.530 (0.041)	-0.150 (0.048)	-0.093 (0.044)	0.179 (0.055)	0.612 (0.058)	0.413 (0.051)
YCS10	1999	42366	0.464 (0.036)	-0.071 (0.041)	-0.001 (0.038)	0.250 (0.052)	0.609 (0.047)	0.513 (0.045)
YCS11	2001	43635	0.454 (0.029)	-0.076 (0.036)	-0.047 (0.034)	0.176 (0.042)	0.521 (0.041)	0.537 (0.036)
NPD1	2002	2495354	0.661 (0.013)	0.052 (0.015)	0.124 (0.012)	0.374 (0.018)	0.647 (0.017)	0.710 (0.014)
NPD2	2003	2551225	0.663 (0.013)	0.087 (0.015)	0.104 (0.013)	0.384 (0.019)	0.659 (0.018)	0.717 (0.014)
					full[b]			
YCS2	1985	24589	0.297 (0.037)	-0.372 (0.043)	-0.075 (0.068)	-0.137 (0.050)	0.429 (0.066)	
YCS3	1986	20669	0.331 (0.043)	-0.373 (0.051)	0.046 (0.085)	-0.140 (0.057)	0.402 (0.070)	
YCS4	1988	19108	0.437 (0.047)	-0.512 (0.056)	-0.339 (0.058)	-0.021 (0.061)	0.435 (0.074)	0.269 (0.082)
YCS5	1990	50711	0.552 (0.036)	-0.291 (0.039)	-0.140 (0.039)	0.107 (0.047)	0.596 (0.057)	0.249 (0.052)
YCS6	1991	80142	0.543 (0.026)	-0.213 (0.029)	-0.165 (0.030)	0.157 (0.035)	0.670 (0.037)	0.362 (0.039)
YCS7	1993	61196	0.493 (0.028)	-0.164 (0.033)	-0.156 (0.033)	0.165 (0.039)	0.631 (0.041)	0.349 (0.041)
YCS8	1995	54665	0.584 (0.031)	-0.123 (0.034)	-0.134 (0.033)	0.215 (0.041)	0.636 (0.041)	0.483 (0.040)
YCS9	1997	34069	0.539 (0.038)	-0.143 (0.042)	-0.082 (0.040)	0.188 (0.051)	0.671 (0.053)	0.411 (0.047)
YCS10	1999	42366	0.468 (0.032)	-0.068 (0.037)	0.005 (0.035)	0.250 (0.046)	0.615 (0.041)	0.519 (0.043)
YCS11	2001	43635	0.477 (0.026)	-0.054 (0.031)	-0.023 (0.029)	0.185 (0.037)	0.552 (0.035)	0.557 (0.033)
NPD1	2002	2495354	0.663 (0.008)	0.051 (0.008)	0.100 (0.008)	0.367 (0.011)	0.670 (0.010)	0.711 (0.010)
NPD2	2003	2551225	0.664 (0.007)	0.085 (0.008)	0.081 (0.008)	0.382 (0.010)	0.688 (0.010)	0.715 (0.010)

* cohort-specific regressions using OLS

[a] girl dummy

[b] raw plus month of birth dummies (except YCS8-10, for which only year of birth is available), single-sex school, girls-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available), school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, 'A' levels, eligibility, special educational needs, type of housing, father's occupation, mother's occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

[c] average points score in English language and literature

[d] points score in maths

[e] points score in science

[f] average points score in history and geography

Table 13: Summary of subject-specific fixed effects regression results, Points Score*

cohort	year	N	English[c]	maths[d]	science[e]	humanities[f]	languages[g]	other[h]
				raw[a]				
YCS6	1991	80142	0.512 (0.024)	-0.244 (0.029)	-0.190 (0.029)	0.138 (0.035)	0.633 (0.036)	0.332 (0.037)
YCS7	1993	61196	0.488 (0.027)	-0.170 (0.032)	-0.164 (0.032)	0.173 (0.038)	0.617 (0.039)	0.349 (0.040)
YCS8	1995	54665	0.541 (0.029)	-0.165 (0.033)	-0.178 (0.031)	0.173 (0.040)	0.602 (0.040)	0.429 (0.037)
YCS9	1997	34069	0.527 (0.035)	-0.154 (0.041)	-0.096 (0.038)	0.184 (0.048)	0.653 (0.051)	0.400 (0.047)
YCS10	1999	42366	0.471 (0.030)	-0.066 (0.034)	0.003 (0.033)	0.247 (0.044)	0.622 (0.039)	0.523 (0.043)
YCS11	2001	43635	0.432 (0.025)	-0.099 (0.031)	-0.072 (0.029)	0.151 (0.038)	0.510 (0.034)	0.505 (0.033)
				full[b]				
YCS6	1991	80142	0.393 (0.031)	-0.364 (0.035)	-0.312 (0.035)	0.007 (0.039)	0.516 (0.041)	0.207 (0.042)
YCS7	1993	61196	0.341 (0.036)	-0.316 (0.040)	-0.307 (0.039)	0.017 (0.044)	0.477 (0.045)	0.192 (0.046)
YCS8	1995	54665	0.419 (0.036)	-0.289 (0.039)	-0.298 (0.038)	0.048 (0.044)	0.479 (0.045)	0.316 (0.042)
YCS9	1997	34069	0.377 (0.043)	-0.307 (0.047)	-0.244 (0.046)	0.031 (0.054)	0.515 (0.056)	0.249 (0.053)
YCS10	1999	42366	0.317 (0.039)	-0.220 (0.042)	-0.148 (0.041)	0.093 (0.049)	0.467 (0.047)	0.371 (0.050)
YCS11	2001	43635	0.345 (0.032)	-0.189 (0.036)	-0.157 (0.035)	0.055 (0.042)	0.419 (0.039)	0.421 (0.038)

* YCS6-11 pooled regressions

[a] girl dummy

[b] raw plus single-sex school, girls-only school, selective school, girl in selective school (except YCS2-5, for which information on single-sex schools is not available), school type, gender mix, school size, pupil-teacher ratio, expenditure, qualified staff, support hours, ‘A’ levels, eligibility, special educational needs, type of housing, father’s occupation, mother’s occupation, single father, single mother, ethnicity (Bangladeshi is dropped for YCS2-4) and regional variables (except YCS2-5, for which the only school-level variable is secondary modern and the only peer group variable is type of housing)

[c] average points score in English language and literature

[d] points score in maths

[e] points score in science

[f] average points score in history and geography

[g] average points score in French and German

[h] average points score in cdt, art/design and business studies

Table 14: Fixed effects regression results for NPD1

	Pooled	FE(i)	FE(s)	FE(l)
English				
gender g	0.214 (0.016)	—	0.217 (0.016)	—
interaction k^3g	0.255 (0.020)	0.255 (0.025)	0.255 (0.020)	0.255 (0.025)
interaction k^4g	0.464 (0.029)	0.466 (0.036)	0.467 (0.029)	0.468 (0.035)
selective school S	1.505 (0.089)	0.250 (0.222)	—	—
interaction gS	-0.168 (0.134)	0.051 (0.277)	-0.445 (0.042)	—
single sex sch Q	0.022 (0.100)	-0.333 (0.182)	—	—
all girls' sch G	0.045 (0.132)	0.600 (0.203)	0.337 (0.027)	0.343 (0.033)
k^3,k^4,1				
corr		0.078	-0.060	0.081
g hetero diff[a]	—	0.185 (0.028)	0.033 (0.016)	—
Maths				
gender g	-0.108 (0.018)	—	-0.108 (0.018)	—
interaction k^3g	0.049 (0.019)	0.049 (0.108)	0.049 (0.020)	0.049 (0.011)
interaction k^4g	0.182 (0.034)	0.187 (0.108)	0.167 (0.034)	0.191 (0.011)
selective school S	1.677 (0.114)	0.000 (0.000)	—	—
interaction gS	-0.117 (0.145)	0.036 (0.300)	-0.130 (0.234)	—
single sex sch Q	-0.088 (0.114)	-0.583 (0.434)	—	—
all girls' sch G	0.179 (0.150)	1.196 (0.300)	0.691 (0.030)	0.705 (0.037)
k^3,k^4,1				
corr		-0.050	-0.022	-0.015
g hetero diff[a]	—	-0.179 (0.036)	-0.011 (0.032)	—
Science				
gender g	-0.071 (0.019)	—	-0.068 (0.019)	—
interaction k^3g	-0.051 (0.021)	-0.051 (0.026)	-0.051 (0.017)	-0.051 (0.026)
interaction k^4g	0.163 (0.038)	0.168 (0.046)	0.168 (0.017)	0.172 (0.022)
selective school S	1.440 (0.120)	-0.250 (0.403)	—	—
interaction gS	-0.071 (0.152)	-0.008 (0.568)	-0.174 (0.037)	—
single sex sch Q	-0.168 (0.114)	0.000 (0.000)	—	—
all girls' sch G	0.205 (0.150)	0.557 (0.497)	0.648 (0.032)	0.652 (0.039)
k^3,k^3,1				
corr		-0.043	-0.057	-0.023
g hetero diff[a]	—	-0.064 (0.033)	-0.009 (0.031)	—

[a] In FE(i), regression of $\widehat{\alpha}_i$ on g_i and constant. In FE(s), regression of $\widehat{\eta}_s$ on g_i and constant.

Figures

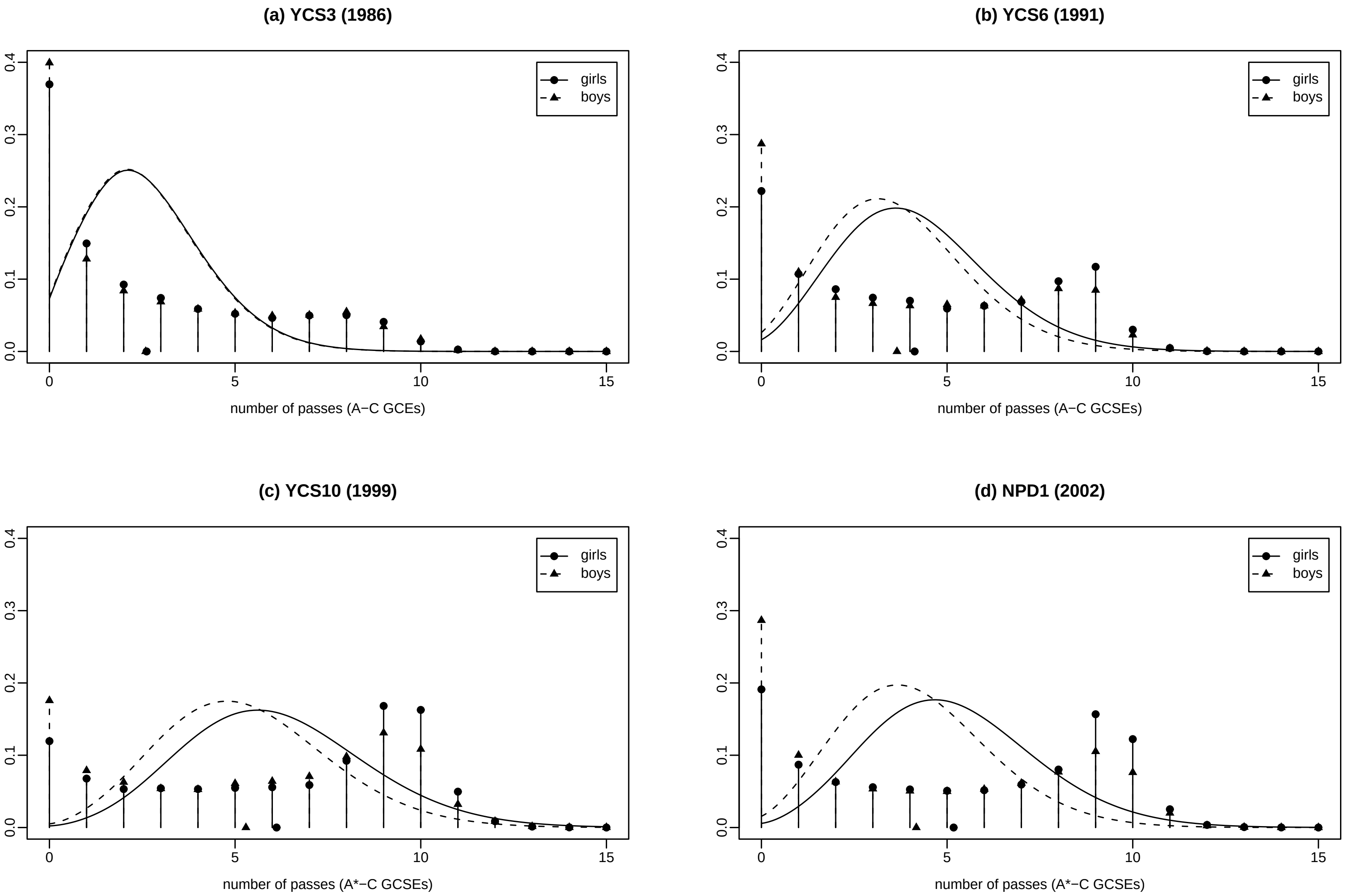


Figure 1: Empirical pdfs and Poisson densities, Number of Passes

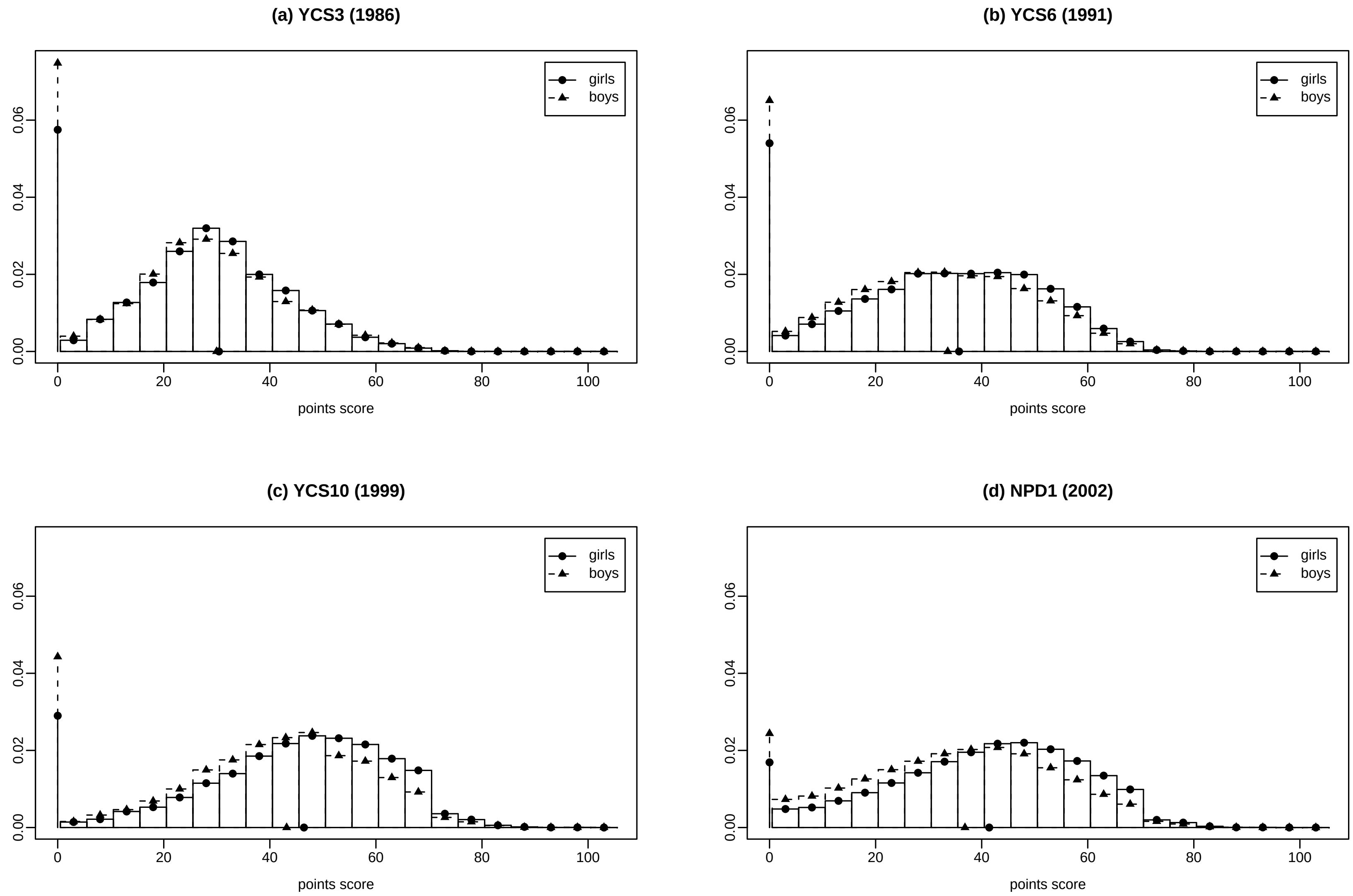


Figure 2: Empirical pdfs, Points Score

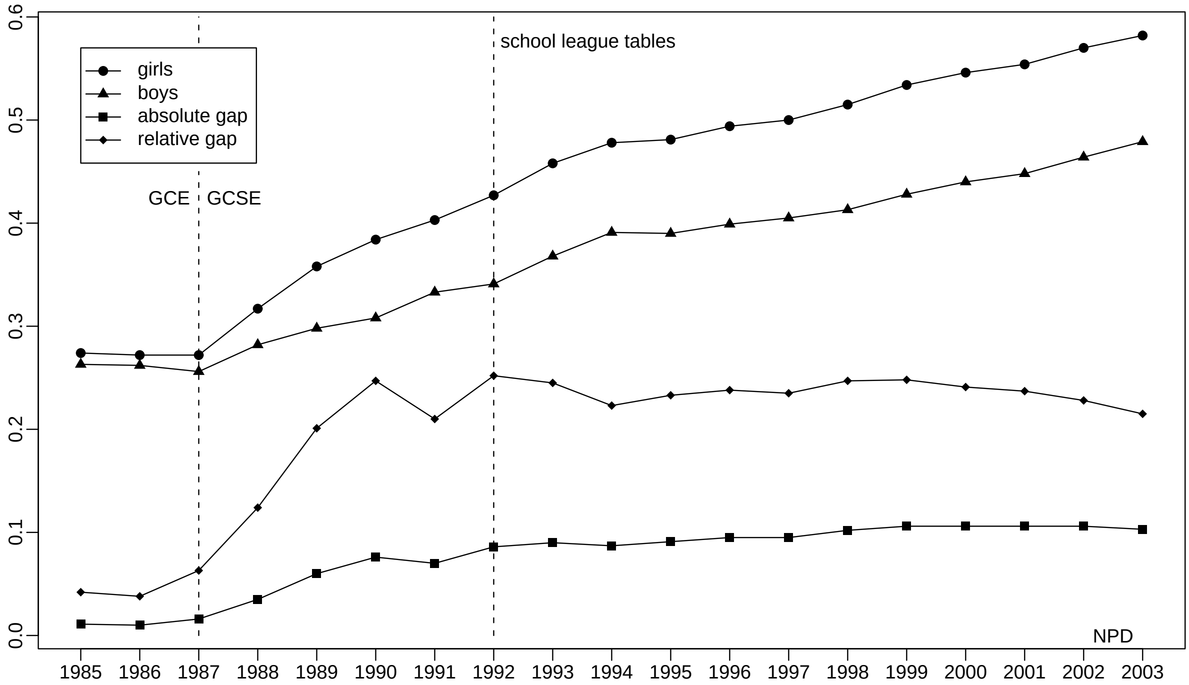
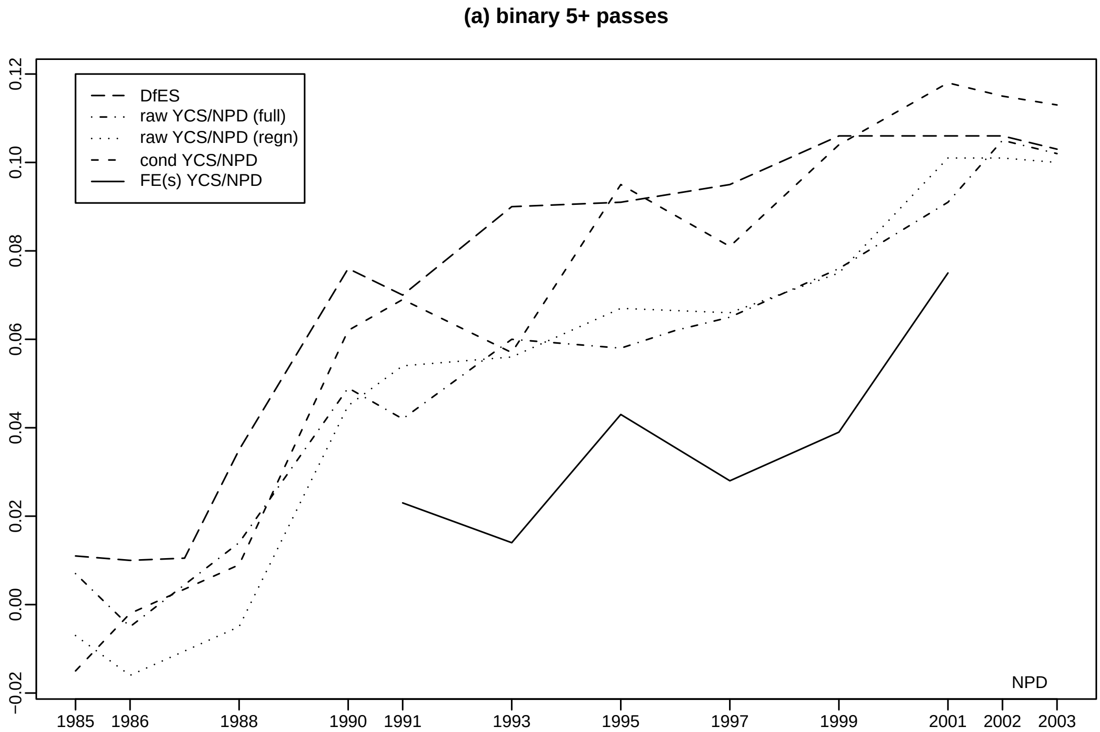


Figure 3: Absolute versus relative gender gaps, Binary 5+ Passes (5+ A*-C GCSEs)



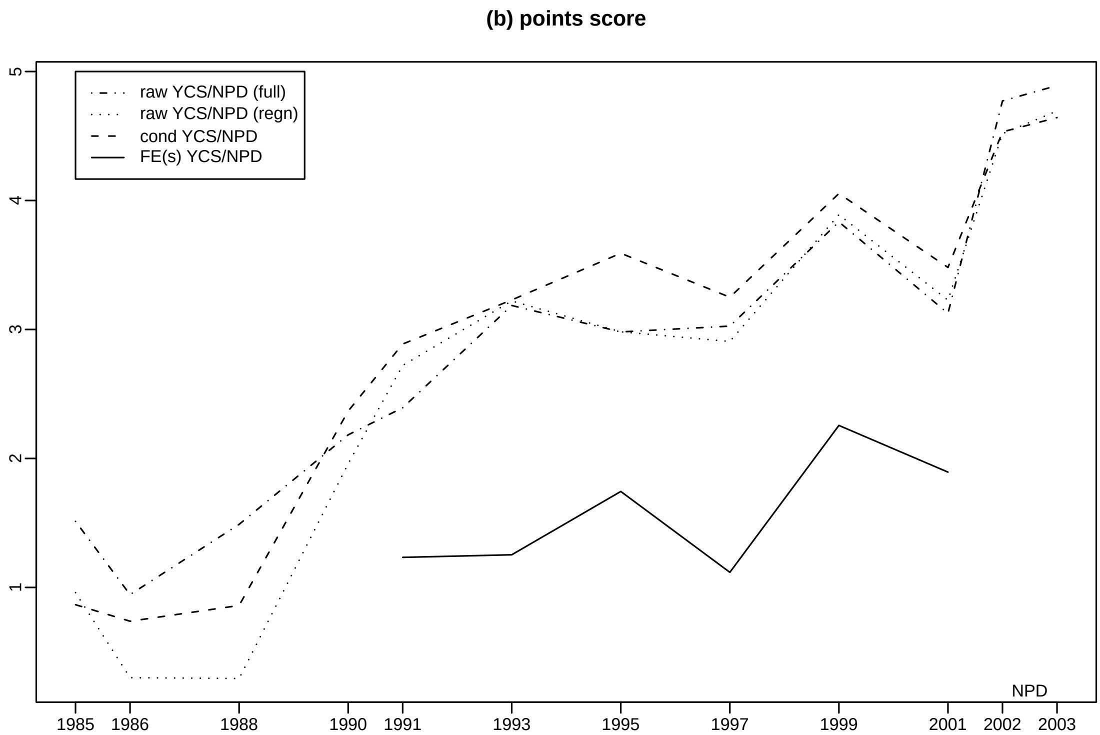
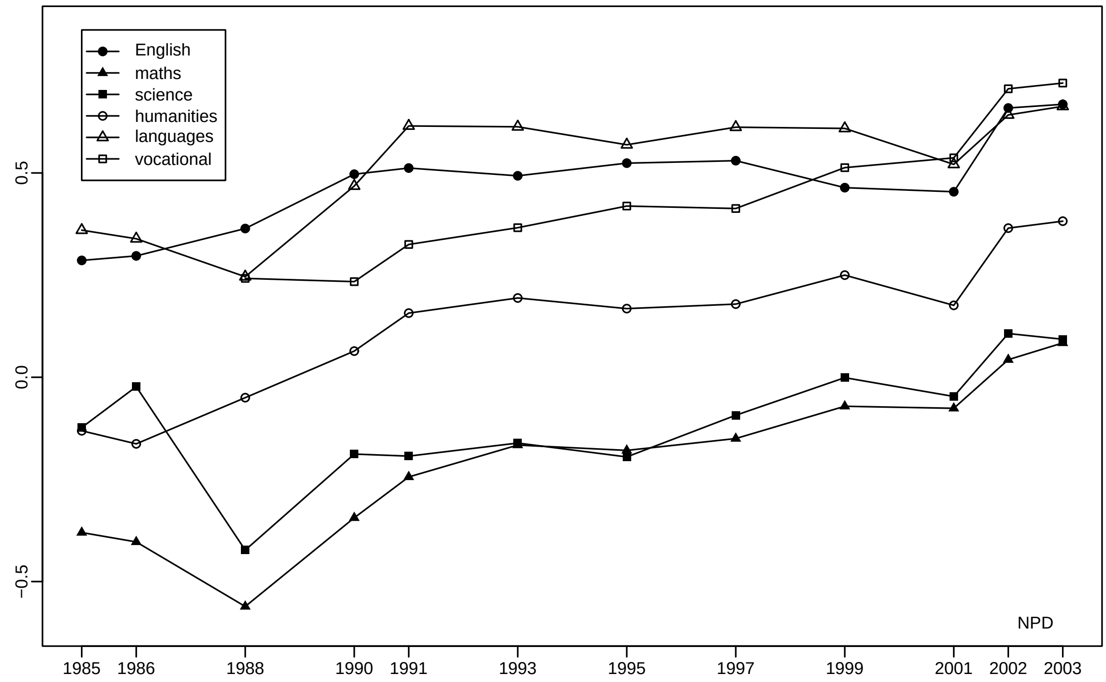


Figure 4: Controlling school unobservables: raw and conditional gender differentials

(a) raw



(b) full, fixed effects

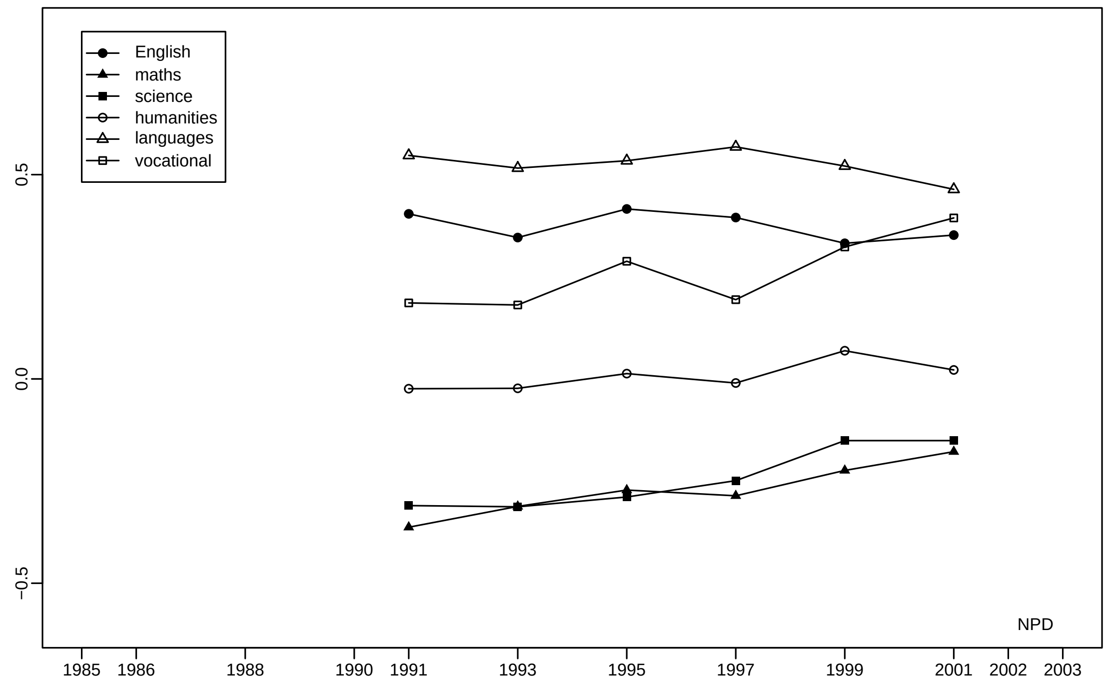


Figure 5: By subject: raw and fixed-effects gender differentials

Table A.1: Notation

i	individual (pupil)
s	school
r	neighbourhood
k	subject
t	cohort
y_i	GCSE attainment (various measures)
y_{ik}	dummy for a GCSE "pass" at subject k
y_{it}	attainment for KS2, KS3, and KS4 (GCSE) (NPD only)
g_i	girl dummy
$\mathbf{x}_i$	fixed characteristics (e.g. family background)
$\mathbf{x}_{s(i)}$	school-level variables (e.g. school size, gender mix)
$\mathbf{x}_{r(i)}$	neighbourhood variables
α_i	unobserved innate ability of individuals (e.g. motivation)
η_s	unobserved school quality (e.g. ethos, discipline)

References

Andrews, M., Schank, T. & Upward, R. (2005*a*), High wage workers and low wage firms: negative assortative matching or statistical artefact? Mimeo, December.

Andrews, M., Schank, T. & Upward, R. (2005*b*), Practical fixed effects estimation methods for the three-way error components model. Mimeo, July.

Atkinson, A. & Wilson, D. (2003), The widening gender gap in English schools, Working Paper No. 99, CMPO, Bristol University.

Bradley, S., Crouchley, R., Millington, J. & Taylor, J. (2000), 'Testing for quasi-market forces in secondary education', *Oxford Bulletin of Economics and Statistics* **62**, 357–90.

Bradley, S. & Taylor, J. (2004), The economics of secondary schooling, *in* G. Johnes & J. Johnes, eds, 'International Handbook of Education Economics', Edward Elgar, Cheltenham.

Burgess, S., McConnell, B., Propper, C. & Wilson, D. (2003), 'Girls rock, boys roll: an analysis of the age 14-16 gender gap in English schools', *Scottish Journal of Political Economy* **51**, 209–29.

Dolton, P. J. & Vignoles, A. (2002), 'Is a broader curriculum better?', *Economics of Education Review* **21**, 415–29.

Dolton, P., Makepeace, G., Hutton, S. & Audas, R. (1999), Making the grade: education, the labour market and young people, Work and Opportunity Series 15, Joseph Rowntree Foundation.

Duncan, G. J. & Dunifon, R. (1998), ''Soft skills' and long run labour market success', *Research in Labour Economics* **17**, 123–49.

Elwood, J. & Murphy, P. (2002), 'Tests, tiers and achievement: gender and performance at age 16 and 14 in England', *European Journal of Education* **37**, 395–416.

Gorard, S., Rees, G. & Salisbury, J. (1999), 'Reappraising the apparent under-achievement of boys at school', *Gender and Education* **11**, 441–54.

Gorard, S., Rees, G. & Salisbury, J. (2001), 'Investigating the patterns of differential attainment of boys and girls at school', *British Educational Research Journal* **27**, 125–39.

Leslie, D. & Drinkwater, S. (1999), 'Staying on in full-time education: reasons for higher participation rates among ethnic minority males and females', *Economica* **66**, 63–78.

Mortimore, P. & Sammons, P. (1994), 'School effectiveness and value added measures', *Assessment in Education* **1**, 315–32.

Murnane, R. J., Willet, J. B. & Levy, F. (1996), 'The growing importance of cognitive skills in wage determination', *The Review of Economics and Statistics,* **99**, 251–66.

Murphy, P. (1999), The role of assessment in differential performance,, *in* J. Salisbury & S. Riddell, eds, 'Gender Policy and Educational Change', Routledge, London.

OFSTED (2003), Boys achievement in secondary schools, HMI 1659, OFSTED Publications Centre, London.

Salisbury, T., Rees, G. & Gorard, S. (1999), 'Accounting for the differential of boys and girls', *School Leadership and Management* **19**, 403–26.

Thomas, S., Sammons, P., Mortimore, P. & Rees, R. (1997), 'Differential secondary school effectiveness: comparing performance of different pupil groups', *British Educational Research Journal* **23**, 451–69.

Tinklin, T. (2003), 'Gender differences and high attainment', *British Educational Research Journal* **29**, 307–325.

Warrington, M. & Younger, M. (1999), 'Perspectives on the gender gap in English secondary schools', *Research Papers in Education* **14**, 51–78.

Warrington, M., Younger, M. & Williams, J. (2000), 'Student attitudes, image and the gender gap', *British Educational Research Journal* **26**, 393–407.

Wong, K., Lam, R. & Ho, L. (2002), 'The effects of schooling on gender differences', *British Educational Research Journal* **28**, 827–43.