Institute of Education

The role of schools in explaining individuals' subject choices at age 14[1]

By Jake Anders**, Morag Henderson**, Vanessa Moulton**, and Alice Sullivan**

Centre for Longitudinal Studies
Working paper 2017/9

CENTRE FOR LONGITUDINAL STUDIES

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

Embargoed to 00.01 hours, 22 August 2017

Contact the author
Jake Anders
Senior Research Associate
Department of Learning and Leadership
UCL Institute of Education
jake.anders@ucl.ac.uk

Jake Anders is funded by the

Nuffield Foundation

** UCL Institute of Education

This working paper was first published in July 2017
by the Centre for Longitudinal Studies,
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institution of Education (IOE). It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ioe.ac.uk.

Embargoed to 00.01 hours, 22 August 2017

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: clsfeedback@ioe.ac.uk

Contents

Acknowledgements

Supported by The Nuffield Foundation under grant EDU/42169. The Nuffield Foundation is an endowed charitable trust that aims to improve social wellbeing in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org. Morag Henderson, Vanessa Moulton and Alice Sullivan's time on this work was supported by the Economic and Social Research Council (Grant number: ES/M008584/1). Thanks to Catherine Dilnot, among others, for helpful comments and suggestions.

Abstract

The subjects that young people study from age 14 onwards may have important consequences for their future academic and labour market outcomes. These decisions are shaped by the schools in which they find themselves. Schools also face constraints of their own. This paper explores the extent to which individuals' decisions are affected by the school they attend and to what extent this is affected by the composition of schools in terms of academic attainment, gender and socioeconomic background. We use multi-level variance decomposition models applied to administrative data on the subjects that young people in mainstream state-funded schools in England study between ages 14 and 16. Our results highlight the important role that constraints on schools play in subject choice decisions. We also note the particular role of attending a non-selective school within a selective schooling district.

Keywords

Subject Choice, Socioeconomic Status, Gender, Schools, Hierarchical Modelling.

Non-technical summary

The subjects that young people study from age 14 onwards may have important consequences for their future academic and labour market outcomes. Choosing the 'wrong' set of options at this point may have long term consequences. However, these decisions are shaped by the schools in which individuals find themselves at this point in time. Furthermore, schools may face constraints of their own. They face significant constraints most obviously from government policy but also, potentially, in responding to what they can offer given the make-up of their student body.

This paper uses administrative data on the subjects that young people in mainstream state-funded schools in England study between ages 14 and 16 to explore the extent to which these are associated with both their own characteristics and the composition of their schools in terms of academic attainment, gender and socio-economic background. We first rank subjects by the average prior academic attainment of pupils who study for each one, providing us with a measure of the 'academic selectivity' of subjects.

Pupils with similar levels of prior attainment or socio-economic status, but who are in schools with higher average levels of these factors, are likely to study a more academically selective curriculum during this period. Individuals in selective or single sex schools tend to study more academically selective subjects. By contrast, individuals in non-selective schools in selective areas tend to study less academically selective subjects. Around one third of variation in subjects studied is explained by which school an individual is it.

These results highlight that we should be sceptical of considering young people's subjects of study purely in terms of 'choice'. They are, at most, constrained choices, potentially both for individuals and for schools. The findings regarding subjects studied in schools within selective local authorities suggest that expanding selective education will increase inequality in and decrease the average level of academic selectivity of subjects that young people study.

1. Introduction

Young people’s subject choices at age 14 may have important consequences for their future academic and labour market outcomes, since they in turn affect the qualifications to which they can easily continue in post-compulsory education. Choosing the ‘wrong’ set of options at this point may have long term consequences (Iannelli, 2013). This is a particularly important issue in an English context, where specialisation of the curriculum occurs earlier than in many other countries (Hodgson and Spours, 2008), and at the height of the education policy of the New Labour government (1997-2010), which promoted diversity and flexibility in the 14-16 curriculum, including in ways that have been criticised for allowing the possibility of ‘gaming’ of performance in school accountability measures (Burgess et al., 2005).

The choices that individuals face are shaped by the schools in which they find themselves at this point in time, just as previous work has found that pupils’ options are restricted depending on where in the country they live (Open Public Services Network, 2015). Schools may not offer certain subjects (Jin et al., 2011) and often guide their pupils towards certain paths (McCrone et al., 2005), for example requiring that a wider set of core subjects be studied, or preventing pupils from taking certain combinations of options. This implies that schools potentially have an important influence in this regard (Woods, 1976). However, schools do not set such requirements in isolation. They face significant constraints most obviously from government policy but also, potentially, in responding to what they can offer given the make-up of their student body. For example, they cannot viably offer an optional subject that only a handful of pupils wish to study. Similarly, the local education market may influence schools’ behaviour (Davies et al., 2003), especially the presence of selective schools.

This paper explores this important issue by assessing the extent to which the subjects that individuals study from age 14-16 are associated, not only with their personal characteristics, but also with the school they attend. It also builds on this, to provide evidence about the extent to which schools’ provision (and, hence, the options open to their pupils) is driven by their composition in terms of prior academic attainment, socioeconomic background and gender mix. These are all factors that previous work has found to influence the subjects that individuals study (Davies et al., 2008), but given Jin et al.’s (2011) observation that the most common reason an individual reports they cannot study a subject is that it is not available, it seems important to explore to what extent individual subject choices are driven by the curriculum offered by schools.

After an initial descriptive exploration, we apply hierarchical regression modelling in order to explore how variance is partitioned between- and within-schools, conditional on other factors. The main results are estimated using administrative data from the National Pupil Database. We also demonstrate the robustness of the results to use richer measures of individuals’ socioeconomic status by replicating our broad findings using survey data from Next Steps (the Longitudinal Study of Young People in England) as far as possible. Overall, our results demonstrate the important role that schools seem to play in subject choice decisions, with significant variation in subjects studied attributable to the school-level. However, they also highlight the potential role of contextual factors for schools, with a significant proportion of this school-level variation being explained by school composition and differences in subjects studied among pupils in non-selective schools in local areas with selective schooling.

The paper proceeds as follows. Section 2 reviews the previous literature on the role of schools in shaping young people's subject choices. Section 3 introduces the datasets used to explore this issue, highlighting the advantages available from using both survey and administrative data to ensure the robustness of the results. Section 4 explores the patterns of variation in subject choices and the extent to which they vary within and between schools. This is developed using hierarchical regression modelling, which is introduced in Section 5; the results of this modelling are reported in Section 6. Finally, Section 7 concludes.

2. Background and research questions

Previous studies that have explored the determinants of subjects studied between ages 14 and 16, have tended to highlight that three important characteristics in explaining subject choices at this age are gender (Bell, 2001; Francis, 2000; Jin et al., 2011; Sullivan et al., 2010), prior attainment (Davies et al., 2008; Jin et al., 2011) and socio-economic background (Davies et al., 2008; Jin et al., 2011).

Bell (2001) considered changes in the uptake of combinations of age 14-16 subjects by gender and prior attainment and how this changed with the introduction of the National Curriculum. Davies et al. (2008) used the 1998 Year 11 Information System (Ye11IS) data to examine the probability of taking GCSEs in optional subjects (specifically: business studies, French, geography, German, history and home economics), finding that 'ability' has the strongest influence on subject choice but for some subjects social class exerts more of an effect than gender. Using a more recent cohort, Jin et al. (2011) find that girls are more likely to study modern foreign language at school and less likely to study all three sciences separately; these associations remain after taking into account prior attainment. Furthermore, those with more educated parents are more likely to study triple science and to stay on in full-time education after Year 11, however these effects are not significant after controlling for prior attainment. Sullivan et al. (2010) also make use of Next Steps data to examine the social structure of the Key Stage Four curriculum in England. They examine how the subjects that young people like or dislike shape the choices that they make and how they are influenced by those around them. In addition, they note a gender difference in vocational subjects and differences by ethnicity for triple science and religious study participation.

Previous work on the importance of subject choice during secondary school has also focussed on specific elements of the decision, for example considering whether young people study science, technology, engineering and maths (STEM)-related subjects (Tripney et al., 2010; Codiroli, 2015). In the case of STEM, this reflects a concern that there is a gender gap in uptake of such subjects, although Codiroli (2015) highlights that this not be the case among individuals from advantaged backgrounds.

Aspects of the role that schools play in shaping subject choice have also been considered. Jin et al. (2011), using Next Steps, document significant variation in the kinds of qualifications offered by different schools. In addition, they note that “19% of pupils were unable to take subjects they would like to study at Key Stage 4” (Jin et al., 2011, p.63), with the most common reason being that their school did not offer the subject (just over 30% of such cases). They identified large differences by school, where some schools offer courses in both academic and vocational choices in Year 10 while others only offer academic courses. Davies et al. (2008) also considered the influence of school context, noting associations between school cohorts and probability of taking subjects, for example the proportion of children who are eligible for free school meals in the school has an effect on the probability of taking certain subjects.

There is a small amount of international evidence on the effects of the sex composition of co-educational schools and classes. Hoxby (2000) uses data on schools in Texas to show that a higher proportion of boys in the class depresses the attainment of both male and female students in both maths and English. Hoxby suggests various possible mechanisms for such peer effects, including classroom disruption and changes in classroom atmosphere. Israeli research also suggests that a high proportion of boys in a year group is linked to worse academic outcomes for both girls and boys (Lavy and Schlosser, 2011). Van Houtte (2004) produces similar findings for Belgium. Proud (2014) uses PLASC/NPD data for the UK, and finds that a higher proportion of girls in the class has a negative effect on boys’ attainment in English, while a higher proportion of girls has a positive effect on both girls’ and boys’ science attainment. Sullivan (2009) found that teenage girls in the 1970s rated their abilities in maths and sciences higher if they went to an all-girls school. Boys on the other hand rated their abilities in English higher if they went to an all-boys school (Sullivan, 2009). Similarly, boys and girls who attended single-sex schools showed increased attainment in gender-atypical subject areas Sullivan et al. (2009), suggesting that single-sex schools may contribute to breaking down gender stereotypes.

It is well established that students attending schools with a high proportion of peers of low social status or low academic ability are at a disadvantage (Coleman et al., 1966; Henderson et al., 1978; Mortimore, 1988; Rutter, 1982; Smith et al., 1989; Willms, 1986). Recent research has suggested that school-SES has no direct effect on individual level attainment except via the academic composition of the school (Marks, 2015). The mechanisms behind school composition effects have not been established empirically. School composition effects may reflect peer group processes (for example if lower-attaining peers are more disruptive). School composition may also influence teachers and the curriculum, as teachers seek to provide a curriculum and pedagogical style which they deem appropriate for the population of the school as a whole (Woods, 1976).

This paper builds on these previous studies by focusing specifically on the role that schools play in explaining subject choice at age 14. We seek to address the following research questions:

- How do the subjects that young people study from ages 14 to 16 vary depending on the school's composition in terms of prior attainment, socioeconomic status and gender?
- What proportion of the variation in young people's subjects of study is between schools, rather than within schools?
- To what extent is the proportion of variation between schools explained by school composition in terms of prior attainment, socioeconomic status and gender?

3. Data

This paper uses data from mainstream English state-funded schools for the academic year 2005-06 available in the National Pupil Database (an administrative dataset owned by the UK's Department for Education) in order to explore these questions. This includes comprehensive academic attainment data from national examinations in England. Rather than self-reports about subjects of study we use the observed information about which GCSEs (or equivalents) young people have entered at age 16. The advantages of administrative data are clear, in that we have information about full cohorts within schools. The NPD also includes some basic data providing a proxy for young people's socio-economic background.

However, this is obviously less fine grained than the background characteristics available in survey data. To test the robustness of the results to this limitation, we replicated the analysis (with the exception of the selective schooling area analyses) using Next Steps (a longitudinal survey that seeks to represent the same population). Full details of this replication are provided in Appendix A. While the results do not replicate exactly (especially regarding school-level relationships), which is perhaps unsurprising given differences in the measurement instruments and the fact that only a small number of students within each school are surveyed, many of the same broad patterns are evident. Where this is not the case, differences are noted and our confidence in these findings is reduced.

We construct a continuous measure of the academic selectivity of the subjects that a pupil studies from age 14-16, based on the prior academic performance of the pupils that choose to study each subject. We assign each subject the average score in Key Stage 3 (KS3) compulsory tests at age 14 of those pupils that report they are studying that subject. KS3 tests are taken roughly contemporaneously with subject choice decisions, so they seem the most appropriate measure to use in this way. The score for a range of subjects is reported in Table 1. We see that those with the highest levels of KS3 attainment are more likely to study subjects such as languages, while those with lower levels are more likely to take applied subjects of various types.

Table 1: Average academic performance at age 14 of pupils studying each GCSE subject ranked in ascending order

Subject	Academic Selectivity Score
Single Science	31.14
Applied Science	31.77
Applied Home Economics	32.99
Other Foreign Lang.	33.85
Applied Business	34.46
Applied Media	34.78
Applied Office	34.82
Other Applied	34.82
Citizenship	35.29
Maths	35.32
English	35.35
Art	35.39
Design Technology	35.48
Drama	35.57
Applied IT	35.58
Double Science	35.64
Religious Education	35.92
Information Technology	36.21
Geography	36.63
History	37.17
French	37.38
Music	37.65
Spanish	37.71
German	38.17
Italian	38.47
Biology	41.40
Chemistry	41.70
Physics	41.75

Notes: Constructed by calculating average point scores in KS3 tests in English, maths and science at age 14 among all individuals who study each subject.

We next convert this into an individual-level, rather than a subject-level, measure. To do so, we sum the top eight most academically selective subjects that each individual studies. A maximum of eight subjects are used to create this measure in order to stop individuals taking a large number of low-selectivity subjects ending up with a high selectivity score. This follows the logic used in the construction of 'capped' GCSE points scores. Thus, individuals who take a combination of academically selective subjects end up with a high score, while individuals who take a combination of less selective subjects are assigned a low score. For ease of interpretation, we standardise this score among the sample used in this paper, so that it has a mean of zero and a standard deviation of one. Henderson et al. (2016) provide further discussion of the development of this measure. Most reassuringly, we note that this method produces similar results to more complex approaches aimed at capturing subject 'difficulty' (Coe et al., 2008).

In addition, we explore the role of schools in explaining the variation in young people making specific sets of decisions that have previously been argued to be important for future outcomes.

Since 2011, the Russell Group (a self-selecting group of highly selective English universities) produces an annual document called "Informed Choices" providing advice on the kind of subject choices that will give young people "the most options" when it comes to accessing such universities (Russell Group, 2013). Although the cohort we analyse predate the publication of this guidance and its primary focus on advice regarding post-16 subjects (highlighting the importance of the 'facilitating' subjects of English Literature, Maths, History, Geography, Languages, Physics, Chemistry and Biology) rather than subject choices at age 14-16, we nevertheless think it useful to analyse which students were already following the spirit of this advice. We analyse whether young people are studying at least three of what might be considered pre-cursors to these post-16 'facilitating' subjects: English, maths, history, geography, languages, physics, chemistry and biology.

Concern has been expressed about the value to pupils of studying 'applied' subjects at GCSE level, especially during the period of our cohort when there were particular worries that schools were using such subjects to 'game' school accountability measures (Burgess et al., 2005). They are certainly among the least academically selective subjects, which is hardly surprising as they were used as alternatives to more academically focused subjects for individuals for whom these seemed less appropriate. We analyse whether individuals study any of these applied subjects.

Partly in response to the concerns raised in the previous paragraph, more recent government policy has sought to incentivise young people to take a combination of subjects known as the English Baccalaureate (EBacc) (made up of English, maths, history or geography, at least two sciences and a modern or ancient foreign language) arguing that this provides young people with the skills they need for the future. Others have countered that this approach may be harmful because the focus on ensuring passes in these subjects may be to the detriment of other subjects. As with Informed Choices, our cohort predates the designation of the EBacc and there was no performance measure incentivising schools to focus on these subjects. Nevertheless, it is useful to analyse which students were already following a curriculum of this type. We analyse whether young people study the full set of these subjects.

Finally, we consider whether young people study all three separate sciences (physics, chemistry, and biology). This full set seems particularly important for progression to STEM subjects, where there has been particular concern about inequality in uptake between males and females.

Given the focus of this paper on the variation between schools the set of independent variables to be used in the model is deliberately parsimonious.

Table 2: Distribution of pupils from each quintile group of KS2 performance and the quintile group of the school's intake as measured by KS2

	Individual						
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Total	Group Size
Q1 (Low)	0.36	0.24	0.19	0.13	0.08	1.00	(0.09)
Q2	0.26	0.23	0.21	0.18	0.12	1.00	(0.17)
Q3	0.20	0.22	0.21	0.20	0.17	1.00	(0.22)
Q4	0.15	0.19	0.22	0.22	0.22	1.00	(0.27)
Q5 (High)	0.07	0.12	0.17	0.25	0.40	1.00	(0.25)
Overall	0.18	0.19	0.20	0.21	0.22	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944.

We include young people's academic attainment at age 11. This age is used since it is at this point that young people sort into secondary schools. The attainment measure at age 11 is based on young people's performance in Key Stage 2 tests in English, maths and science. We standardise this variable to be a Z-score (i.e. mean of zero, standard deviation of one) to aid interpretation. The distribution of pupils by quintile group of their KS2 performance and the quintile group of their school's KS2 intake is reported in Table 2. While schools with low KS2 intakes have, by definition, larger numbers of pupils whose KS2 performance is in the bottom quintile group, there are still individuals in such schools with high levels of KS2 performance.

Table 3: Distribution of pupils from each quintile group of SES and the quintile group of the school's SES intake

	Individual						
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Total	Group Size
Q1 (Low)	0.61	0.24	0.09	0.04	0.02	1.00	(0.10)
Q2	0.28	0.32	0.21	0.12	0.07	1.00	(0.16)
Q3	0.13	0.25	0.26	0.21	0.15	1.00	(0.21)
Q4	0.06	0.15	0.25	0.29	0.25	1.00	(0.26)
Q5 (High)	0.02	0.07	0.16	0.29	0.46	1.00	(0.27)
Overall	0.16	0.18	0.21	0.22	0.23	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944.

The NPD includes two proxies for individuals' socioeconomic status, namely whether they are eligible for free school meals (FSM) and the deprivation status of their neighbourhood (Chowdry et al., 2013). These are combined using principal components analysis with a polyserial correlation matrix (Olsson, 1979; Kolenikov and Angeles, 2009) to construct a single index of SES (alternative methods, such as factor analysis, yield very similar results). This explains roughly three quarters of the variation in the two individual measures. Again, we ensure that this has a mean of zero and a standard deviation of one. There is a similar pattern of the SES distribution within and between schools as that described above regarding prior attainment (Table 3).

Table 4: Distribution of pupils by gender and group of the school's gender distribution

	Individual			
School	Female	Male	Overall	Group Size

All Female	1.00		1.00	(0.08)
Mainly Female	0.58	0.42	1.00	(0.18)
Gender Balanced	0.51	0.49	1.00	(0.60)
Mainly Male	0.41	0.59	1.00	(0.08)
All Male		1.00	1.00	(0.05)
Overall	0.52	0.48	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944. School gender distribution groups are as follows: All female: Prop. male = 0.00; Mainly female: 0.00 < Prop. male 0.45; Gender balanced: 0.45 ≤ Prop. male ≤ 0.55; Mainly male: 0.55 < Prop. male < 1.00; All male: Prop. male = 1.00. Group sizes may not quite sum to 1 due to rounding.

On gender, we find that 60% of pupils are in broadly gender balanced schools, while 13% are in single gender schools (8% female-only and 5% male only), 18% are in mainly female schools, and 8% are in mainly male schools.

4 Subject choice, individual characteristics and school composition

We begin with a descriptive exploration of differences in young people's subject choices depending not only on their personal characteristics, but also on the composition of their school in terms of prior attainment, socioeconomic status, and gender. In the interests of space, this descriptive exploration focuses only on our subject academic selectivity measure; we broaden our focus to other specific subject choices in Section 5.

Average academic subject selectivity is positively associated with both an individual's prior attainment and the average prior attainment of their schoolmates (Table 5). However, this is not just because the school-level association picks the average change in individuals' prior attainment within the school. Individuals with low prior attainment, in schools with high average prior attainment, are more likely to study more academically selective subjects than their counterparts in lower average prior attainment schools. In other words, student intake appears to matter for improving opportunities for students to study more academically selective subjects.

Table 5: Average academic subject selectivity of individuals by their KS2 performance and the average KS2 performance of their schoolmates

	Individual

School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Overall
Q1 (Low)	-0.77	-0.50	-0.33	-0.15	0.04	-0.59
Q2	-0.66	-0.36	-0.18	-0.01	0.20	-0.24
Q3	-0.54	-0.22	-0.03	0.15	0.37	-0.02
Q4	-0.50	-0.16	0.03	0.23	0.44	0.21
Q5 (High)	-0.38	-0.07	0.17	0.39	0.66	0.49
Overall	-0.48	-0.27	-0.07	0.05	0.35	0.00

Notes: Cells report average standardised capped academic subject selectivity. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944.

There is a similar overall pattern when it comes to socioeconomic status, perhaps in part because there is a correlation (*r*=0.28) between the measures of SES and performance at KS2. On an individual level, SES is, reassuringly, less predictive of the academic selectivity of subjects studied than young people's prior attainment. Notably, the SES intake of a school appears just as important a predictor as the individual-level indicator.

Table 6: Average academic subject selectivity of individuals by their SES background and the average SES of their schoolmates

	Individual					
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Overall
Q1 (Low)	-0.51	-0.33	-0.20	-0.26	-0.14	-0.44
Q2	-0.49	-0.28	-0.18	-0.09	0.00	-0.17
Q3	-0.38	-0.19	-0.06	0.06	0.08	0.03
Q4	-0.25	0.00	0.12	0.19	0.29	0.15
Q5 (High)	-0.21	0.02	0.19	0.25	0.33	0.26
Overall	-0.42	-0.27	-0.09	0.14	0.25	0.00

Notes: Cells report average standardised capped academic subject selectivity. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944.

Academic selectivity of subjects studied also varies somewhat by gender, with girls studying a set of courses with a slightly higher academic selectivity, on average than boys (Table 7). However, in this case, the selectivity of subjects studied varies much more by the gender composition of their school. Most noticeably, pupils in single sex schools study a set of subjects of significantly higher academic selectivity than those in mixed schools.

Table 7: Average academic subject selectivity of individuals by their gender and the gender composition of schoolmates

	Individual		
School	Female	Male	Overall
All Female	0.31	-	0.31
Mainly Female	-0.06	-0.19	-0.11
Gender Balanced	0.02	-0.09	-0.03
Mainly Male	-0.00	-0.06	-0.04
All Male	-	0.36	0.36
Overall	0.05	-0.06	0.00

Notes: Cells report average standardised capped academic subject selectivity. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 542,944.

However, it is important to recognise that the analysis above only attempts to explore one aspect at a time of the association between young people's characteristics, schools' composition and the subjects studied by its pupils. There are reasons to think that some of the associations are due to other confounding factors, especially in relatively small groups such as single-sex schools. In order to alleviate this issue, and to allow us to explore the importance of these factors for specific subject choices in a manageable way, we now turn to hierarchical regression modelling.

5 Hierarchical regression modelling

In order to explore the variation in subject choice in a more formal framework, attempting to decompose the influence of prior attainment, SES and gender, we fit hierarchical regression models of the outcomes of interest. Hierarchical models (also known as multi-level models or regression models with random effects) explicitly partition the variance in the model between units and within units (in this case schools). Marks (2006) uses a similar technique when looking at within- and between-school variation in academic performance.

Hierarchical regression models of the following form are estimated. For continuous outcome variable (index of subject selectivity):

$$y_{ij} = \beta_0 + \beta X_{ij}' + \gamma \bar{X}_j' + \eta_j + \varepsilon_{ij} \quad (1)$$

For dichotomous outcome variables:

$$ln\left(\frac{y_{ij}}{1-y_{ij}}\right) = \beta_0 + \beta X_{ij}' + \gamma \bar{X}_j' + \eta_j + \varepsilon_{ij} \quad (2)$$

where y is the outcome variable, i indicates the individual, j indicates the school which they attend, X represents a vector of individual-level regressors, $\bar{X}$ represents a vector of the school-level averages of the individual-level regressors, and β and γ represent fixed-effects coefficients. Finally, there are two error terms: η represents a normally distributed school-level random intercept, and ε represents an individual-level error term (normally distributed in the case of models with a continuous outcome and with a logistic distribution in the case of models with a dichotomous outcome).

Models including random effects make assumptions about relationships in the data, which are typically unlikely to be strictly justified. First, it is assumed that the school-level effects are normally distributed. Second, it is assumed that there is no correlation between the individual-level and the school-level error terms. However, one strategy for relaxing this latter assumption is modelling this relationship by including group-level means of the regressors (Mundlak, 1978). Given their substantive interest, these are being included in the model in any case. As such, this gives us increased confidence in the estimated relationships between the individual-level covariates and the outcome variables.

Given our interest in the role of schools, an important element of the results is the proportion of variance explained simply by the school which an individual attends i.e. not the proportion explained by within-school variation or the proportion explained by school-level variation in the make-up of the pupils. This is known as the intra-cluster correlation (ICC or ϱ) and is calculated from the hierarchical models as follows:

$$\varrho = \frac{var(\eta)}{var(\eta)+var(\varepsilon)} \qquad (3)$$

Covariates included in the model are added in a sequential manner as follows. First, an empty model (M0) is estimated. This performs the important function of providing a baseline unconditional intra-cluster correlation against which the conditional intra-cluster correlations in later models may be compared.

In the first model including covariates (M1), we add an individual's KS2 standardised score (Z-Score) as well as the school's average KS2 Z-Score. This provides results on the conditional association between young people's prior attainment after controlling for the school's performance. It aims to capture not only whether individuals with higher prior attainment study more academically selective subjects but also whether there is an additional increase in the average subject mix of individuals in schools with an intake of more highly attaining pupils.

The next two models (M2 and M3) adds in covariates related to gender and SES, respectively; as with M1 these capture both whether these individual-level characteristics are relevant for subjects studied and whether the school context in these terms has additional predictive power. In M2 we include a dummy variable for an individual's gender. Given the findings of Section 4, we allow for a more flexible relationship between the school's gender balance and subject choice a simple linear relationship. We add dummy variables that categorise schools into gender-balanced (our baseline), mainly male, mainly female, all-male, and all-female. M3 adds the standardised index of young people's SES, along with the school's average value of this index. This provides evidence on whether there is an independent association between SES and subject choice once prior attainment has been held constant.

Our final model (M4) looks specifically at a possible constraint on schools. We include an indicator for whether individuals are in a school located in a Local Authority (LA) in which more than 5% of pupils attend selective (grammar) schools and for whether they are in a selective school themselves. While much of the dynamic of being in a school with high prior attainment will be captured through the covariates introduced in M1, we are particularly interested to see whether being in a non-selective school within an area in which selective schools are present affects the subjects studied.

6 Results

The results tables focus on the conditional associations between individual-level characteristics and subject choices, school-level compositions and subject choices, and the proportion of variance explained at the school level conditional on included covariates (ϱ). For models of the continuous index of academic selectivity we report the regression coefficients, which are interpreted as the expected change in the subject selectivity index in standard deviations.

Table 8: Variation in subject selectivity

	M0	M1	M2	M3	M4
KS2 Z-Score		0.321	0.322	0.308	0.308
		(76.93)***	(77.21)***	(76.10)***	(76.10)***
School KS2 Z-Score		0.420	0.379	0.193	0.172
		(17.48)***	(14.23)***	(6.30)***	(4.14)***
Male			-0.127	-0.130	-0.130
			(-24.13)***	(-24.77)***	(-24.77)***
All-male school			0.156	0.270	0.278
			(3.91)***	(6.37)***	(6.35)***
Mainly-male school			0.0302	0.0998	0.105
			(0.83)	(2.77)**	(2.91)**
Mainly-female school			-0.0722	-0.0611	-0.0604
			(-2.60)	(-2.27)*	(-2.24)*
All-female school			0.103	0.230	0.239
			(3.22)**	(6.70)***	(6.75)***
SES Z-Score				0.107	0.107
				(39.75)***	(39.75)***
School SES Z-Score				0.124	0.129
				(4.90)***	(5.01)***
>5% in LA in selective schools					-0.0603
					(-2.00)*
Selective school					0.0659
					(1.22)
Dep. Var. Mean	0.00	0.00	0.00	0.00	0.00
ρ	0.32	0.27	0.27	0.26	0.26
N	344148	344148	344148	344148	344148

Notes: Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Omitted categories are female and being in a gender-balanced school. Reporting regression coefficients. t statistics in parentheses. Stars indicate statistical significance: + *p*<0.10, * *p*<0.05, ** *p*<0.01, *** *p*<0.001. Cond. ϱ reports intra-cluster correlation coefficient from reported model.

We consider the results for the subject selectivity score first (Table 8). In M0, in which there are no regressors, the only estimate to be interpreted is the proportion of variance between schools rather than within schools, which is estimated to be 0.32. This suggests that approximately two thirds of the variation in individuals' subject selectivity is between different pupils within the same school. After conditioning on school intake in terms of prior attainment (M1) this proportion of variation explained by schools is reduced to 27% in models M1; this is only reduced marginally further after conditioning on gender or socioeconomic status in M2 and M3.

Turning to associations between characteristics and subject choices, on adding controls for prior attainment at age 11 (M1), we find that a one standard deviation increase in an individual's prior attainment is associated with roughly a 0.3 of a standard deviation change in subject selectivity score. The schools' intake attainment profile is associated with subject selectivity even more strongly, with a one standard deviation change in school intake being associated with 0.4 of a standard deviation increase in the average academic selectivity of the subjects its pupils study.

Adding gender and schools' gender balance to the model (M2) makes little difference to the previously described relationships between prior attainment and subject selectivity. Nevertheless, we see a statistically-significant association between an individual's gender and the academic selectivity of the subjects they study, with girls on average studying a set of subjects that are 0.13 standard deviations more academically selective than their comparable male peers. School-level gender effects are particularly marked when comparing single-sex schools to gender balanced ones with both kinds of single-sex school entering pupils for more academically selective subjects.

Adding socioeconomic status to the model (M3) we find that both individuals' SES and the average SES in their school are positively associated with the academic selectivity of the subjects they study; these seem of similar qualitative importance. Furthermore, adding the controls for SES reduces the association between school's KS2 intake and subject choice markedly, although it remains an important factor.

Finally, in M4, we consider selective schools and those in local areas where academically selective schools take a significant proportion of individuals (specifically, local authority areas where at least 5% of the population attend a selective school). Including indicators for whether an individual is in a selective area and whether they are in a selective school in the model makes little difference to the other covariates. However, pupils in a school in a selective area study for a statistically significantly less academically selective mix of subjects. Those actually in selective schools see an approximately offsetting effect. This suggests that this result is driven by pupils in non-selective schools in selective areas studying less academically selective subjects, even after individual and school-level factors have been taken into account.

Table 9: Odds ratios from logistic regression models of probability that individuals study specific combinations of subjects at age 14

	3 Facil.	Applied	EBacc	Triple Sci.
KS2 Z-Score	2.022	0.861	2.116	1.060
	(70.76)***	(-12.19)***	(75.10)***	(27.81)***
School KS2 Z-Score	2.288	0.426	2.551	1.034
	(7.44)***	(-4.55)***	(8.14)***	(2.99)**
Male	0.969	0.710	0.918	1.016
	(-1.95)+	(-13.31)***	(-5.49)***	(13.93)***
All-male school	1.490	0.403	1.773	1.099
	(3.47)***	(-3.59)***	(4.35)***	(4.87)***
Mainly-male school	1.115	1.023	1.020	1.019
	(1.13)	(0.13	(0.20)	(2.16)*
Mainly-female school	0.838	0.948	0.820	1.004
	(-2.58)**	(-0.44)	(-2.70)**	(0.69)
All-female school	2.398	0.477	2.063	0.988
	(7.65)***	(-3.71)***	(7.53)***	(-1.02)
SES Z-Score	1.279	0.993	1.253	1.004
	(36.00)	(-1.02)	(31.44)***	(6.62)***
School SES Z-Score	1.209	0.857	1.581	1.001
	(3.29)***	(-1.59)	(6.95)***	(0.17)
>5% in LA in selective schools	0.824	1.077	0.761	0.987
	(-2.52)*	(0.55)	(-3.36)***	(-1.93)+
Selective school	3.909	0.0758	1.235	1.168
	(7.07)***	(-7.27)***	(1.26)	(6.09)***
Dep. Var. Mean	0.81	0.31	0.37	0.08
Uncond. ρ	0.40	0.67	0.42	0.29
Cond. ρ	0.30	0.64	0.32	0.23
N	344148	344148	344148	344148

Notes: Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Omitted categories are female and being in a gender-balanced school. Reporting odds ratios (i.e. exponentiated coefficients from the logistic regression model). t statistics in parentheses. Stars indicate statistical significance: + *p*<0.10, * *p*<0.05, ** *p*<0.01, *** *p*<0.001. Uncond. ϱ reports intra-cluster correlation coefficient from model of dependent variable including no covariates; Cond. ϱ reports intra-cluster correlation coefficient from reported model. Column titles refer to model dependent variables, as follows: 3 Facil. = Individual studies for at least three 'facilitating' subjects; Applied = Individual studies for at least one 'applied subject'; EBacc = Individual studies a full set of subjects that would go to be required for eligibility for English Baccalaureate; Triple Sci. = Individual studies for separate qualifications in Physics, Chemistry and Biology.

Turning to our binary models of individuals studying particular combinations of subjects (Table 9), we restrict our discussion to highlighting important differences between the general subject selectivity model and these more specific models. We only report the models including the full set of covariates we consider (M4). We also report the intra-cluster correlation from this model (Cond. ϱ) and from the model not including any covariates (Uncond. ϱ) for comparison. For these models we report odds ratios (exponentiated logistic regression coefficients), which may be interpreted as the expected change in the odds of studying the relevant set of subjects.

We start our discussion with whether individuals study at least three ‘facilitating’ subjects, something that we find 81% of the sample to do. As robustness checks we have also run models of whether individuals study at least four or five ‘facilitating’ subjects and find broadly similar results at these alternative margins. The pattern of significant coefficients is broadly similar to the model of our academic selectivity measure, although we cannot compare the magnitudes with the selectivity model due to its linear outcome. One standard deviation higher prior attainment by an individual at age 11 is associated with twice the odds of studying at least three facilitating subjects. Much less dramatically, individuals with one standard deviation higher SES have 28% increased odds of studying three or more facilitating subjects, while males have just slightly reduced odds of doing so.

There is a large and statistically significant association with school intake in terms of prior attainment: a one standard deviation increase in the intake of the school on this measure is associated with a more than doubling of the odds of studying at least three facilitating subjects. The overall importance of schools, as measured by the intra-cluster correlation, is similar to that in the overall model, both unconditionally and conditionally on the set of controls. There is also evidence of increased odds of studying these subjects in single sex schools of either type, especially all-female schools. There is also an association with the SES intake of the school of a similar magnitude of that seen for the individual-level SES association. Even after accounting for these other characteristics, those in areas with selective schools see a significant reduction the odds of studying at least three facilitating subjects, while those actually in selective schools have almost four times the odds of doing so.

Whether individuals study any applied subjects, something that is true of 31% of the sample, has a contrasting set of associations. There is a negative relationship between individuals’ prior attainment at age 11 and the odds of studying any applied subjects: individuals with a one standard deviation higher KS2 score have 14% lower odds of studying an applied subject. There is no significant difference between individuals’ SES and odds of studying an applied subject, although a negative relationship was observed in our survey data model, perhaps due to the stronger measurement of SES in that setting. One area of similarity between the overall selectivity model and the applied subjects model is that male students have reduced odds of studying any applied subjects. Likewise, individuals in all-male schools see one-quarter the odds of studying applied subjects compared to their peers in gender-balanced schools; a similar (but not quite as large) reduction is also evident in all-female schools. Again, the differences associated with selective schooling are stark: those in selective schools are over 90% lower odds of studying any applied subjects than peers in non-selective schools.

Another difference that stands out is the much greater value of intra-cluster correlation, compared to the general selectivity score model. Less than half of the variance is between individuals within the same schools, perhaps suggesting schools play a particularly large role in whether or not individuals study such subjects. Furthermore, this influence of schools seems divorced from the makeup of schools in this case, since the conditional intra-cluster correlation is only marginally lower than the unconditional one.

We next turn to the model of whether individuals are studying the full set of subjects that would be necessary to be eligible for the English Baccalaureate (EBacc) (if they were in a later cohort and also go on to reach the required standard in the qualifications). Just over a third (37%) of the sample meet this criteria. Individuals with one standard deviation higher KS2 scores have over twice the odds of studying the EBacc subjects, while those in schools with one standard deviation higher KS2 intake see their odds of studying all EBacc subjects increase by over two and a half times. Individuals with one standard deviation higher SES have approximately 25% increased odds; those in schools with a one standard deviation higher intake in terms of SES see their odds increase by just over 50%. These are broadly similar patterns to those seen in the models of general subject selectivity in terms of the significance of individuals' prior attainment and SES, and schools' average intake in terms of these two characteristics.

The individual-level gender difference is much smaller than for the overall selectivity score but, once again, there is a much increased chance of taking EBacc subjects in all-male or all-female schools. Again, the presence of selective schooling in an area appears to be associated with a reduced probability of taking a full set of subjects, except in selective schools themselves where the opposite is true. Finally, schools also appear to play a bigger role in explaining whether young people study EBacc subjects than they did in explaining our general selectivity score, with a conditional intra-cluster correlation of 0.37.

We next consider variation in whether individuals study three separate sciences. This is the smallest group that we consider, at only 8% of the sample. Once again, those in selective schools are more likely to study triple sciences, even after taking into account other characteristics in the model. Students in all-male schools see 10% increased odds of studying this subject combination, relative to those in gender balanced schools while, in contrast to other subject combinations we have considered, there is no significant change in the odds among pupils in all-female schools. Even holding these influences of school constant, boys are a little more likely to study three separate sciences than their female peers. This is in contrast to findings for overall academic subject selectivity, where male pupils were more likely to study less academically selective subjects.

Conditional on the covariates in the model, and in contrast to the other subject combination models, whether individuals study triple sciences sees a lower proportion of variation explained by schools to that estimated in overall selectivity score, with a conditional intra-cluster correlation of 0.22.

7 Conclusions

This paper has explored the issue of variation in young people's subject choices by assessing the extent to which individuals' decisions are associated with the school they attend. However, it also goes beyond this, to explore the importance of constraints on schools, especially how schools' provision (and, hence, the options open to their pupils) is shaped by their composition in terms of academic attainment, socio-economic background and gender mix.

This work includes no analysis of later outcomes following from studying more academically selective subjects or, indeed, the particular combinations of subjects on which we have focussed. This paper does not offer evidence on whether there are particular benefits from particular combinations of subject choice. It is, however, something to which we plan to return in future work. As such, our conclusions offer no judgement on whether we should consider the differences that we document to be positive or negative, only that they exist.

Our results replicate previous findings that young people's prior attainment, socio-economic background, and gender are all associated with the subjects they study at age 14-16. However, our novel contribution is to consider the separate associations between the composition of their school in these terms and subjects studied. We find that individuals in schools with more advantaged intakes are more likely to study more academically selective subjects, even after conditioning on individuals' own SES. Individuals' prior attainment is associated with studying more academically selective subjects as, again, is the prior attainment of the school more generally. Boys are less likely to study academically selective subjects than their female counterparts, although there is also evidence of higher subject selectivity in single-sex schools of either type.

Overall, schools explain about a third of the variation in the academic selectivity of the subjects that young people study; once we take into account the demographics of the school this is reduced to closer to a quarter. We suggest two potential mechanisms. First, schools may try to offer a curriculum which they deem appropriate for the socio-economic composition of the school; in other words, schools may deliberately take school SES into account when choosing the subjects offered. Second, we know that schools serving poor children face difficulties in recruiting and retaining highly qualified staff (Lupton, 2005; Lupton and Thrupp, 2012; Ofsted, 2013) particularly in shortage areas such as languages and sciences. This may constrain the curriculum that schools with disadvantaged intakes are able to offer. The strength of the school SES effect is surprising, and has potential policy implications, as it suggests an additional pathway through which school SES influences academic outcomes for pupils. The negative association between being in a local authority in which a significant proportion of pupils attend selective schools and the academic selectivity of subjects that individuals study raises the possibility that local education markets (for pupils and teachers) also impose constraints on the subjects that schools may viably offer.

There are some important differences when we consider some examples of whether individuals study for specific sets of subjects that it has been argued may be important for future academic outcomes. The odds of studying any applied subjects is lower among pupils in schools with a higher prior attaining intake. While it is the case that in most single sex schools pupils are more likely to study a more academically selective set of subjects, for studying triple sciences this is only true in all-male schools and not all-female ones.

Overall, this paper has highlighted the important role that schools seem to play in many subject choice decisions, with significant variation in subjects studied attributable to the school-level. However, it also highlights that, in many cases, what we might see as schools' actions are, in fact, strongly associated with contextual factors, as captured by their intake in terms of prior attainment, gender and socioeconomic status; one plausible pathway for this would be that intake imposes constraints on what schools can viably offer. The socio-economic composition of the school is a powerful predictor of individual choices in our models, even controlling for the academic composition of the school. In fact, the SES composition of the school has a similar strength of relationship with subject selectivity as the individual's family SES.

These findings have a number of implications. First, they highlight that we should be sceptical of considering young people's subjects of study purely in terms of 'choice' (Woods, 1976). They are, at most, constrained choices, potentially both for individuals and for schools. Second, the findings regarding subjects studied in schools within selective local authorities suggest that expanding selective education will increase inequality in, and decrease the average level of, the academic selectivity.

References

Anders, J. (2012). Using the Longitudinal Study of Young People in England (LSYPE)/“Next Steps” for research into Higher Education access. DoQSS Working Paper 12-13, Institute of Education, University of London.

Bell, J. F. (2001). Patterns of Subject Uptake and Examination Entry 1984-1997. *Educational Studies*, 27(2):201–219. doi:10.1080/03055690120050428.

Breen, R. and Yaish, M. (2006). Testing the Breen-Goldthorpe Model of Educational Decision-Making. In Morgan, S. L., Grusky, D. B., and Fields, G. S., editors, *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, pages 232–258. Stanford University Press.

Burgess, S., Propper, C., Slater, H., and Wilson, D. (2005). Who wins and who loses from school accountability? the distribution of educational gain in english secondary schools. CMPO Working Paper 05/128, Centre for Market and Public Organisation, University of Bristol.

Chevalier, A. (2004). Parental education and childs education: a natural experiment. CEE Discussion Paper CEE DP 40, Centre for the Economics of Education, London School of Economics and Political Science.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A*, 176(2):431–457. doi:10.1111/j.1467-985X.2012.01043.x.

Codiroli, N. (2015). Inequalities in students’ choice of stem subjects. CLS Working Paper 2015/6, Centre for Longitudinal Studies, UCL Institute of Education.

Coe, R., Searle, J., Barmby, P., Jones, K., and Higgins, S. (2008). Relative difficulty of examinations in different subjects. CEM Centre Summary Report, Centre for Evaluation and Monitoring, Durham University.

Coleman, J. S., Campbell, E. Q., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. L. (1966). *Equality of Educational Opportunity*. National Center for Educational Statistics, U.S. Department of Health, Education, and Welfare, Washington, DC.

Davies, P., Adnett, N., and Turnbull, A. (2003). Market forces and diversity: Some evidence from the 14-19 curriculum. *Journal of Curriculum Studies*, 35(4):479–498. doi:10.1080/00220270305519.

Davies, P., Telhaj, S., Hutton, D., Adnett, N., and Coe, R. (2008). Socioeconomic background, gender and subject choice in secondary schooling. *Educational Research*, 50(3):235–248. doi:10.1080/00131880802309358.

Ermisch, J. and Pronzato, C. (2010). Causal effects of parents’ education on children’s education. ISER Working Paper 2010-16, Institute of Social and Economic Research, University of Essex.

Francis, B. (2000). The Gendered Subject: Students’ subject preferences and discussions of gender and subject ability. *Oxford Review of Education*, 26(1):35–48. doi:10.1080/030549800103845.

Goldthorpe, J. H. and McKnight, A. (2004). The Economic Basis of Social Class. CASEpaper 80, Centre for Analysis of Social Exclusion, London School of Economics.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, 65(6):1541–1545. doi:10.2307/1131279.

Havari, E. and Savegnago, M. (2014). The causal effect of parents' schooling on children's schooling in europe. a new iv approach. CEIS Tor Vergata Research Paper Vol. 12, Issue 5, No. 315, Cenre for Economic and International Studies Tor Vergata.

Henderson, M., Sullivan, A., and Anders, J. (2016). Social class, gender and ethnic differences in subjects taken at age 14. CLS Working Paper, Centre for Longitudinal Studies, UCL Institute of Education.

Henderson, V., Mieszkowski, P., and Sauvageau, Y. (1978). Peer group effects and educational production functions. *Journal of Public Economics*, 10(1):97–106. doi:10.1016/0047-2727(78)90007-5.

Hodgson, A. and Spours, K. (2008). *14-19 Education and Training: Curriculum, Qualifications and Organisation*. Sage.

Hoxby, C. (2000). Peer effects in the classroom: learning from gender and race variation. NBER Working Paper 7867, National Bureau of Economic Research.

Iannelli, C. (2013). The role of the school curriculum in social mobility. *British Journal of Sociology of Education*, 34(5-6):907–928. doi:10.1080/01425692.2013.816031.

Jenkins, S. P. and Schluter, C. (2002). The effect of family income during childhood on later-life attainment: evidence from Germany. IZA Discussion Paper 604, Institute for the Study of Labor.

Jin, W., Muriel, A., and Sibieta, L. (2011). Subject and course choices at ages 14 and 16 amongst young people in England: insights from behavioural economics. DfE Research Report DFE-RR160, Department for Education.

Kolenikov, S. and Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*, 55(1):128–165. doi:10.1111/j.1475-4991.2008.00309.x.

Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33. doi:10.1257/app.3.2.1.

Lupton, R. (2005). Social justice and school improvement: improving the quality of schooling in the poorest neighbourhoods. *British Educational Research Journal*, 31(5):589–604. doi:10.1080/01411920500240759.

Lupton, R. and Thrupp, M. (2012). Headteachersâ€™ readings of and responses to disadvantaged contexts: evidence from english primary schools. *British Educational Research Journal*, pages 1–20. doi:10.1080/01411926.2012.683771.

Marks, G. N. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? evidence from 30 countries. *Educational Research*, 48(1):21–40. doi:10.1080/00131880500498396.

Marks, G. N. (2015). Are school-ses effects statistical artefacts? evidence from longitudinal population data. *Oxford Review of Education*, 41(1):122–144. doi:10.1080/03054985.2015.1006613.

McCrone, T., Morris, M., and Walker, M. (2005). Pupil Choice at Key Stage 3 - Literature Review. NFER Report, National Foundation for Educational Research.

Mortimore, P. (1988). *School Matters*. University of California Press.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85. doi:10.2307/1913646.

Ofsted (2013). The report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2013/14: Schools. Annual Report, Ofsted.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460. doi:10.1007/BF02296207.

Open Public Services Network (2015). Lack of options: how a pupil's academic choices are affected by where they live. OPSN Report, Open Public Services Network.

Proud, S. (2014). Girl power? an analysis of peer effects using exogenous changes in the gender make-up of the peer group. *Regional and Sectoral Economic Studies*, 14(3):5–18.

Rose, D. and Pevalin, D. J. (2001). The National Statistics Socio-economic Classification: Unifying Official and Sociological Approaches to the Conceptualisation and Measurement of Social Class. ISER Working Paper 2001-4, Institute for Social and Economic Research, University of Essex.

Russell Group (2013). Informed Choices: A Russell Group guide to making decisions about post-16 education. Russell Group guide.

Rutter, M. (1982). *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*. Harvard University Press.

Smith, D. J., Tomlinson, S., Bonnerjea, L., Hogarth, T., and Thomes, H. (1989). *The School Effect: A Study of Multi-Racial Comprehensives*. Policy Studies Institute, London.

Sullivan, A. (2009). Academic selfâ€ concept, gender and singleâ€ sex schooling. *British Educational Research Journal*, 35(2):259–288. doi:10.1080/01411920802042960.

Sullivan, A., Joshi, H., and Leonard, D. (2009). Single-sex schooling and academic attainment at school and through the lifecourse. *American Educational Research Journal*, 47(1):6–36. doi:10.3102/0002831209350106.

Sullivan, A., Zimdars, A., and Heath, A. (2010). The social structure of the 14â€“16 curriculum in England. *International Studies in Sociology of Education*, 20(1):5–21. doi:10.1080/09620211003655572.

Tripney, J., Newman, M., Bangpan, M., Niza, C., MacKintosh, M., and Sinclair, J. (2010). Subject Choice in STEM: Factors Influencing Young People (aged 14-19) in Education. A systematic review of the UK literature. Wellcome Trust Report, EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

Van Houtte, M. (2004). Gender context of the school and study culture, or how the presence of girls affects the achievement of boys. *Educational Studies*, 30(4):409–423. doi:10.1080/0305569042000310336.

Willms, J. D. (1986). Social class segregation and its relationship to pupils' examination results in scotland. *American Sociological Review*, 51(2):224–241.

Woods, P. (1976). The myth of subject choice. *British Journal of Sociology*, 27(2):130–149.

Appendix

Replication in Next Steps

Where possible, we repeat the analysis from this paper using data from Next Steps (formerly known as the Longitudinal Study of Young People in England). This has the disadvantage of sometimes only having a small number of individuals within each school, which is likely to bias downwards the proportion of variance which the models attribute to school. However, it has much richer data regarding young people's socioeconomic background.

Next Steps follows a cohort of young people born in 1989-90 from age 14 through to age 20. The survey has a clustered design based around schools, so that young people are randomly selected for inclusion within randomly selected schools (albeit with some oversampling). It includes annual interviews throughout with the young people themselves, interviews with their parents (for the first four years), and linked administrative data about young people's academic attainment (from the National Pupil Database, discussed above). Using the responses from the parental questionnaires it provides high quality data on young people's socioeconomic background, based on questions about family income, parental education, and occupational status. In addition, in includes self-reported information on subjects that young people are studying at age 14.

Next Steps includes a rich set of data with which to measure young people's socio-economic status (SES), including household income, parental education, and parental occupational status, all of which are important in measuring SES (Hauser, 1994). Household income is measured at each wave between 1 and 4. As previous research has suggested 'permanent' income (rather than transitory income) has a much larger effect on young people's educational outcomes (Jenkins and Schluter, 2002, p.2), an approximation of the household's equivalised 'permanent' income is made by averaging across these four measures and dividing by the square root of household size. Previous work suggests that Next Steps underestimates household income to some extent, relative to social surveys where it is a major focus (Anders, 2012). Parental education also captures an important aspect of socioeconomic status, with one explanation for this being that it "may alter the 'productivity' of [parents'] time investments in children" (Ermisch and Pronzato, 2010, p.1). Whatever the explanation, a number of studies have found evidence of a causal impact of parents education on children's outcomes (Ermisch and Pronzato, 2010; Chevalier, 2004; Havari and Savegnago, 2014), making it an important factor to take into account. Social class is seen by sociologists as a key element of an individual's SES (Goldthorpe and McKnight, 2004), in particular as "young people (and their families) have, as their major educational goal, the acquisition of a level of education that will allow them to attain a class position at least as good as that of their family of origin" (Breen and Yaish, 2006, p.232). Parents' occupational status is recorded in Next

Steps using the National Statistics SocioEconomic Classification (NS-SEC), which was designed to capture social class differences between the different occupational types (Rose and Pevalin, 2001). The above measures of household income, parental education, and parental occupational status are combined (using principal components analysis with a polychoric correlation matrix[1] (Olsson, 1979; Kolenikov and Angeles, 2009)) to construct a single index of SES. This explains roughly three quarters of the variation in the three individual measures, but provides a broader measure of family circumstances than any one measure would provide. Again, we ensure that this has a mean of zero and a standard deviation of one.

We also construct the SES measure using the same variables as those available in the NPD, to allow a closer replication of the results in the main body of the paper. The results using this SES measure (and its school-level average) are reported in model N3 (rather than M3). Due to lack of area-level indicators, it is not possible to replicate model M4 using data from Next Steps. Otherwise, all models follow the same specification as those reported in Section 5.

[1] Alternative methods, such as factor analysis, yield very similar results.

A.1 Data

Table 10: Average academic performance at age 14 of pupils studying each subject in ascending order

Subject	Academic Selectivity Score
Applied Hospitality	29.28
Applied Leisure	29.91
Applied Health	30.45
Applied Manufacturing	30.55
Applied Art	31.36
Other Applied	32.12
Applied Science	32.62
Other Foreign Lang.	33.31
Applied Business	33.31
Citizenship	33.48
Art	33.62
Drama	33.67
Physical Education	33.78
English	33.87
Maths	33.87
Information Technology	34.15
Personal Social Health Ed.	34.55
Religious Education	34.67
Applied IT	34.75
Music	34.95
Geography	35.23
History	35.66
Biology	35.67
Chemistry	35.69
Physics	35.82
French	36.27
Spanish	36.31
Italian	36.65
Statistics	36.87
German	37.46

Notes: Constructed by calculating average point scores in KS3 tests in English, maths and science at age 14 among all individuals who study each subject.

Table 11: Distribution of pupils from each quintile group of KS2 performance and the quintile group of the school's intake as measured by KS2

	Individual						
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Total	Group Size
Q1 (Low)	0.40	0.24	0.18	0.12	0.06	1.00	(0.17)
Q2	0.24	0.25	0.20	0.18	0.12	1.00	(0.21)
Q3	0.17	0.22	0.23	0.21	0.17	1.00	(0.21)
Q4	0.12	0.18	0.23	0.24	0.22	1.00	(0.20)
Q5 (High)	0.05	0.10	0.17	0.25	0.42	1.00	(0.21)
Overall	0.19	0.20	0.20	0.20	0.20	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465.

Table 12: Distribution of pupils from each quintile group of SES and the quintile group of the school's SES intake

	Individual						
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Total	Group Size
Q1 (Low)	0.48	0.26	0.16	0.08	0.02	1.00	(0.17)
Q2	0.24	0.29	0.23	0.17	0.07	1.00	(0.21)
Q3	0.16	0.23	0.25	0.22	0.15	1.00	(0.22)
Q4	0.10	0.17	0.22	0.27	0.24	1.00	(0.24)
Q5 (High)	0.04	0.10	0.18	0.28	0.41	1.00	(0.17)
Overall	0.20	0.21	0.21	0.20	0.17	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465.

Table 13: Distribution of pupils by gender and group of the school's gender distribution

	Individual			
School	Female	Male	Overall	Group Size
All Female	1.00		1.00	(0.05)
Mainly Female	0.60	0.40	1.00	(0.17)
Gender Balanced	0.50	0.50	1.00	(0.43)
Mainly Male	0.41	0.59	1.00	(0.30)
All Male		1.00	1.00	(0.05)
Overall	0.49	0.51	1.00	(1.00)

Notes: Reporting row proportions, except for final column which reports column proportions. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465. School gender distribution groups are as follows: All female: Prop. male = 0.00; Mainly female: 0.00 < Prop. male < 0.45; Gender balanced: 0.45 ≤ Prop. male ≤ 0.55; Mainly male: 0.55 < Prop. male < 1.00; All male: Prop. male = 1.00.

A.2 Descriptives

Table 14: Average academic subject selectivity of individuals by their KS2 performance and the average KS2 performance of their schoolmates

	Individual					
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Overall
Q1 (Low)	-0.53	-0.37	-0.27	-0.01	0.12	-0.48
Q2	-0.52	-0.20	-0.09	0.05	0.19	-0.18
Q3	-0.47	-0.17	0.05	0.22	0.33	-0.00
Q4	-0.33	-0.09	0.06	0.27	0.45	0.20
Q5 (High)	-0.20	0.11	0.20	0.33	0.63	0.45
Overall	-0.34	-0.17	-0.00	0.12	0.39	0.00

Notes: Cells report average standardised capped academic subject selectivity. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465.

Table 15: Average academic subject selectivity of individuals by their SES background and the average SES of their schoolmates

	Individual					
School	Q1 (Low)	Q2	Q3	Q4	Q5 (High)	Overall
Q1 (Low)	-0.38	-0.28	-0.18	-0.13	-0.10	-0.31
Q2	-0.31	-0.14	-0.15	-0.04	0.20	-0.12
Q3	-0.30	-0.19	-0.03	0.08	0.16	-0.01
Q4	-0.15	0.10	0.11	0.25	0.39	0.16
Q5 (High)	0.00	0.14	0.19	0.40	0.49	0.35
Overall	-0.30	-0.14	-0.06	0.19	0.36	0.00

Notes: Cells report average standardised capped academic subject selectivity. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465.

Table 16: Average academic subject selectivity of individuals by their gender and the gender composition of schoolmates

	Individual		
School	Female	Male	Overall
All Female	0.20	-	0.20
Mainly Female	0.03	-0.05	-0.00
Gender Balanced	0.05	-0.01	0.02
Mainly Male	-0.02	-0.17	-0.10
All Male	-	0.24	0.24
Overall	0.05	-0.05	0.00

Notes: Cells report average standardised capped academic subject selectivity. Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Sample size = 12,465.

A.3 Results

Table 17: Variation in subject selectivity

	M0	M1	M2	M3	N3
KS2 Z-Score		0.260	0.258	0.238	0.255
		(20.90)***	(20.81)***	(19.01)***	(19.73)***
School KS2 Z-Score		0.145	0.129	-0.0746	0.0450
		(1.85)+	(1.61)	(-0.87)	(0.54)
Male			-0.0751	-0.0753	-0.0617
			(-3.84)***	(-3.86)***	(-3.10)**
All-male school			0.0825	0.0808	0.110
			(0.83)	(0.86)	(1.09)
Mainly-male school			-0.0290	-0.0176	0.00174
			(-0.55)	(-0.34)	(0.03)
Mainly-female school			-0.226	-0.0108	-0.0171
			(-0.37)	(-0.18)	(-0.29)
All-female school			0.106	0.126	0.166
			(1.75)+	(1.99)*	(2.49)*
SES Z-Score				0.0651	
				(6.11)***	
School SES Z-Score				0.303	
				(4.70)***	
SES Z-Score (NPD)					0.0498
					(3.48)***
School SES Z-Score (NPD)					0.109
					(2.64)**
Dep. Var. Mean	0.00	0.00	0.00	0.00	0.00
ρ	0.26	0.24	0.24	0.23	0.23
N	12050	12050	12050	12050	11137

Notes: Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Omitted categories are female and being in a gender-balanced school. Reporting regression coefficients. t statistics in parentheses. Stars indicate statistical significance: + *p*<0.10, * *p*<0.05, ** *p*<0.01, *** *p*<0.001. Cond. ρ reports intra-cluster correlation coefficient from reported model.

Table 18: Odds ratios from logistic regression models of probability that individuals study specific combinations of subjects at age 14

	3 Facil.	Applied	EBacc	Triple Sci.
KS2 Z-Score	1.863	0.710	2.338	1.015
	(14.57)***	(-10.13)***	(19.31)***	(3.17)**
School KS2 Z-Score	0.823	1.435	1.421	1.058
	(-0.83)	(2.24)*	(2.06)*	(3.70)***
Male	0.847	0.817	0.951	1.018
	(-2.22)*	(-3.58)***	(-0.81)	(2.40)*
All-male school	1.162	0.513	1.754	1.201
	(0.42)	(-3.10)**	(2.55)*	(3.26)**
Mainly-male school	1.010	0.972	0.902	0.995
	(0.07)	(-0.25)	(0.81)	(-0.34)
Mainly-female school	1.181	0.945	1.037	1.015
	(0.97)	(-0.45)	(0.25)	(0.78)
All-female school	1.982	0.601	1.447	0.984
	(2.25)*	(-2.56)*	(2.11)*	(-0.66)
SES Z-Score	1.223	0.787	1.284	1.006
	(4.50)***	(-7.96)***	(7.70)***	(1.44)
School SES Z-Score	2.607	0.564	2.372	1.059
	(5.07)***	(-4.17)***	(5.53)***	(3.31)***
Dep. Var. Mean	0.84	0.47	0.26	0.14
Uncond. ρ	0.33	0.22	0.33	0.19
Cond. ρ	0.28	0.20	0.22	0.15
N	12050	12057	12050	12050

Notes: Weighted using Next Steps Wave 2 non-response and attrition weights. Sample: Young people at state-schools with valid data on subject choices, gender, SES, and KS2 performance. Omitted categories are female and being in a gender-balanced school. Reporting odds ratios (i.e. exponentiated coefficients from the logistic regression model). t statistics in parentheses. Stars indicate statistical significance: + *p*<0.10, * *p*<0.05, ** *p*<0.01, *** *p*<0.001. Uncond. ϱ reports intra-cluster correlation coefficient from model of dependent variable including no covariates; Cond. ϱ reports intra-cluster correlation coefficient from reported model.