

How does the choice of A-level subjects vary with students' socio-economic status in English state schools?

Catherine Dilnot

Work carried out at:

UCL Institute of Education, Department of Social Science, 20 Bedford Way, London WC1H 0AL

Correspondence address:

Oxford Brookes University, Department of Accounting, Finance and Economics, Wheatley Campus, Oxford OX33 1HX

Email address: cdilnot@brookes.ac.uk

Sponsorship

ESRC doctoral training centre grant number ES/J500021/1

Abstract

The reasons why students from lower socio-economic groups are under-represented at highly selective universities are not entirely understood, but evidence suggests that part of the gap may be a consequence of differential choice of A-levels by social background. The Russell Group of universities has since 2011 published guidance on subject choices, describing some A-levels as 'facilitating' in that their choice keeps the largest number of Russell Group degree courses open to potential applicants. This study uses National Pupil Database data from three recent cohorts of English state school students taking at least three A-levels, and a taxonomy of all 96 A-levels certified for English students in 2014/15. Large differentials in subject choice by social background are found, particularly for facilitating subjects but also for subjects considered 'less suitable' by Russell Group universities. Linear probability models show that these differentials substantially disappear when GCSE attainment and subject choices at age 14+ are taken into account. Closing the choice gap at A-level is likely therefore to depend on reducing differentials in attainment and subject choice by social background at GCSE. The introduction of the eBacc may help with the GCSE subject choice element

Introduction

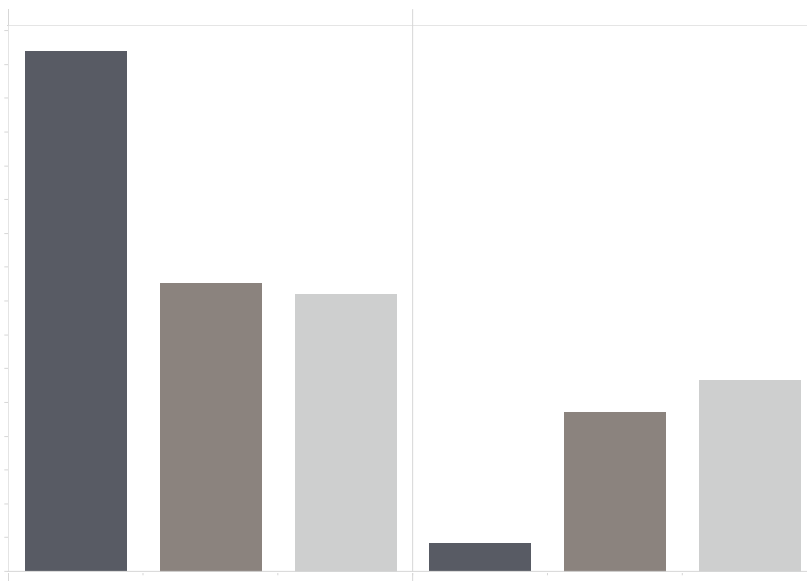
Positive wage returns to an undergraduate degree persist, despite significant recent increases in the number of graduates (Blanden & Macmillan, 2014). In an intergenerationally mobile society, admission to university would depend principally on the efforts and abilities of the individual rather than the socio-economic status (SES) of their parents (Blanden, Gregg, & Macmillan, 2007; Crawford, Johnson, Machin, & Vignoles, 2011). We also know that returns to a degree vary considerably by type of university (Britton, Dearden, Shephard, & Vignoles, 2016; Chevalier & Conlon, 2003; Hussain, McNally, & Telhaj, 2009) and by degree subject choice (Walker & Zhu, 2011, Britton et al. (2016)), so if social mobility is to increase, it is important that those from less privileged backgrounds are able to attend the types of institutions and take the subjects conferring higher returns. There is clear evidence that those from lower SES backgrounds and state schools rather than private schools remain under-represented at university generally and at highly selective universities in particular (Anders, 2012; Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2013; Sullivan, Parsons, Wiggins, Heath, & Green, 2014). The raw participation gap by SES does seem to be reducing for university generally over the last ten years, but more slowly for highly selective universities, and is increasing for the top quintile by achievement at Key Stage 5 (A-level) (Crawford, 2012).

What is it about coming from a lower SES background that means students are less likely to go to a highly selective university? The main barrier seems to be lower school attainment (Anders, 2012; Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2013; Crawford, Macmillan, & Vignoles, 2015; Marcenaro-Gutierrez, Galindo-Rueda, & Vignoles, 2007) but even after prior attainment is taken into account some effect of SES on highly selective university attendance remains (Anders, 2012; Crawford, 2012), suggesting that other factors may also be important. One such factor is A-level subject choice; applicants to highly selective universities may apply with good grades, but in the 'wrong' subjects (Russell Group, 2012). Helping able lower SES students make the 'right' choices might be a relatively easy way to increase their chances of attending a highly selective university.

Working out which are the 'wrong' A-level subjects is not straightforward. There were 96 separately certified A-levels available for English students to take in 2014/15. The Russell Group of 24 large, highly selective universities has since 2011 published annual guidance entitled 'Informed Choices' (Russell Group, 2011) on A-levels it considers helpful for admission to its member institutions. It suggests a list of 'facilitating' subjects (maths, sciences, modern and classical languages, English literature, history and geography). Its argument is that the more of these subjects chosen at age 16+, the more course options will be available at Russell Group universities when students make their applications in their final year (year 13) or afterwards. Counting all languages separately, facilitating subjects account for 33 of the 96 possible A-levels available for teaching in 2014/15. There is a lack of centralised information about the remaining subjects. I have therefore produced a taxonomy, based on the published preferences of Russell Group universities derived from 'Informed Choices', published lists of five Russell Group universities which make public statements about the general acceptability of a range of A-levels, and admissions pages for Russell Group courses in subjects related to A-levels, for which the A-level might plausibly provide useful preparation (Dilnot, 2015). This

taxonomy further categorises non-facilitating subjects as ‘useful’, of ‘more limited suitability’ and ‘less effective preparation’ in the context of highly selective university admission. My analysis of all students with three A-levels at English mainstream schools and colleges (including private schools) entering UK universities between 2010 and 2012 using linked National Pupil Database (NPD) and Higher Education Statistics Agency (HESA) data does indeed suggest significant differences in proportions of subjects from these categories held by those attending Russell Group and non-Russell Group institutions, and not attending a British university as shown in Figure 1. Noting the difference in uptake of facilitating and non-facilitating subjects between private and state school pupils, the previous Government introduced the achievement of high grade A-levels in at least two facilitating subjects as a national social mobility indicator (Office of the Deputy Prime Minister, 2013) and a school level performance measure (Department for Education, 2014).

Figure 1: Proportions of three A-level cohort of students by A-level pattern held and university status



Evidence suggests that the participation gap by socio-economic status at British universities is reduced once A-level choices are taken into account (Boliver, 2013; Chowdry et al., 2013). This begs the question of why SES might affect subject choices. Human capital theory suggests that individuals make educational investment decisions to maximise their future productivity, subject to constraints which differ by SES (Becker & Tomes, 1986). Although choosing one A-level subject rather than another results in no extra financial cost to students, long run factors to do with increased family resources helping to foster better cognitive and non-cognitive skills (Carneiro & Heckman, 2002) mean that students from different social backgrounds are differentially constrained in their choices. The first differential constraint may arise from the effect of prior attainment, for example through schools requiring a threshold GCSE score before allowing an A-level to be chosen. Second, students from different backgrounds are sorted into different schools (Burgess & Briggs, 2010), and will have different choice sets of A-levels available to them, as well as differences in expectations made of them, and of guidance received. Third, there may be information differentials between students of different backgrounds which lead to choices that prevent more deprived students from fulfilling their

academic ambitions (Jackson & Jonsson, 2013). Whitty, Hayton, and Tang (2015) suggest that the experiences and cultural capital of students from lower SES backgrounds are likely to affect their capacity to navigate educational pathways, which may be particularly problematic in the context of elite university entry. One of the ways in which such students may not 'know the ropes' (p44) is the necessity to choose appropriate A-level subjects that will keep options open, probably some considerable time before decisions about particular universities and courses are made.

I contribute to the literature by using my A-level taxonomy to investigate the largely unresearched relationship between SES and choice of A-levels, using linked administrative data from the National Pupil Database (NPD) for three cohorts of students taking A-levels in England from 2010 to 2012, and taking into account school differences, using first single level then fixed effects models. Illumination of the pathways through which SES status is linked with subject choice will help schools provide appropriate information, advice and guidance (IAG) to students making subject choices, and will provide evidence to university admissions teams considering A-level subject course requirements, so contributing to a reduction in inequality of elite university entry. I proceed in this paper by discussing the literature on the relationship between subject choice and social background, followed by setting out the methods and data. The last section includes a discussion of results, some conclusions and next steps.

The relationship between A-level subject choice and SES

The literature on subject choices at age 14+ has covered a wide range of GCSE subjects (for example Davies, Telhaj, Hutton, Adnett, and Coe (2008) and Iannelli (2013)) whilst that on A-level choice to date has largely concentrated on the uptake of STEM (Science, Technology, Engineering, Mathematics) subjects, partly because of the perceived economic need for a workforce trained for the predicted increase in STEM based occupations, and also because of concern over the lack of diversity in those choosing STEM subjects at A-level. Much of this work is concerned with the gender gap in subject choice, but there is also evidence of a relationship with SES. Gorard, See and Smith (2008), in a systematic review of patterns of science participation from the literature and in an analysis of NPD data (Gorard and See, 2009), find students eligible for free school meals significantly less likely than their more privileged peers to take maths and science at A-level. Gill and Bell (2013) find a small reduction in the probability of taking A-level physics for pupils living in areas with low car ownership (although not for other measures of deprivation) once prior attainment is controlled for, and other studies find relationships between the uptake of A-level maths and SES (Cheng, Payne, & Witherspoon, 1995; Sharp, Hutchison, Davis, & Keys, 1996).

Vidal Rodeiro (2007) examines reasons for subject choice across all A-level subjects and finds that reasons for making particular subject choices, and the subjects chosen both relate to parents' occupational class. The children of higher managers and lower managers/professionals are significantly less likely to take at least two science subjects than the children of higher professionals, but the children of higher managers are more likely to take the business related subjects of accounting, business and economics. Manual workers' children

are significantly less likely to take a foreign language A-level, but no parental occupation effect is observed for taking two 'less effective preparation' subjects. No theoretically driven classification of subjects in terms of efficacy of subjects in highly selective university admission was used in this analysis; instead content based groupings of science, language, business, vocational, technologies, creative arts, established humanities and newer humanities were used, together with the 'less effective preparation' list that was then in use by Cambridge University (but has since been withdrawn). The study was of a random stratified sample of 60 schools, with a response rate of 40% from the students within the schools, and did not aim to make estimates generalizable to the population. More recently, a longitudinal study of 3000 students followed since age 3 finds bright but disadvantaged students much less likely to take at least one facilitating A-level subject than their bright but more advantaged peers (Toth, Sammons, & Sylva, 2015). Using NPD data for 2014, Gill (2015b) provides *prima facie* evidence of gaps in the take up of individual A-level subjects and in numbers of facilitating subjects by social background (measured using the Income Deprivation Affecting Children Index) and school type. I contribute to the literature in this area by using NPD data to analyse the choices by social background of all students taking A-levels in three recent cohorts, taking into account school and individual level characteristics.

Literature on the choices made by students in other countries is difficult to translate generally to the English context as it tends to concentrate on the effect of tracked systems rather than choices at age 16, but provides potentially generalizable insights. In a study in the Netherlands, Van de Werfhorst (2002) finds children from working class backgrounds likely at age 12 to choose technical and commerce related subjects available within the vocational rather than general school track. Such subjects are classified as 'less suitable' for highly selective university entry in the UK context, despite their more obvious links to careers than subjects in other categories. As Davies, Qiu, and Davies (2014) suggest from work on university aspirations, cultural capital can help make sense of labour market information, and students from less privileged backgrounds might be less willing to take subjects not obviously linked to a career because they and their parents lack the knowledge that access to a career may not necessarily be via the most obvious route, and that choices maximising their entry chances to a highly selective university might be better for their long term career goals than taking an A-level in a particular career-related subject.

Indications therefore exist in the literature that there are social class effects in uptake of facilitating subjects. The evidence on other subjects is mixed and requires further analysis; there does seem to be an effect for some such subjects (for example accounting and business studies) but not for all of the Cambridge less effective preparation list taken together, once academic attainment is controlled for (Vidal Rodeiro, 2007). The full extent of the pathways by which SES might have an effect on subject choice is as yet unclear. I therefore ask the following questions.

Research questions

1. Are there differences in the proportions of English state school students choosing facilitating or 'less suitable' A-levels by social background, before adjusting for any other characteristics?

2. How much of these differences is explained by prior attainment and choice of GCSE subjects?
3. Is any remaining difference in subject choice by social background accounted for by the observed characteristics of schools, into which students of similar social background tend to be selected?
4. If both observed and unobserved characteristics of schools (such as school ethos and quality of IAG) are taken into account, are there still differences in subject choice by social background?

Methods

I use the taxonomy to investigate the relationship between social background and choice of A-levels for students at English state schools and 6th form and further education colleges as SES data is largely unavailable for students at private schools. A-level outcomes are based on my classification of A-levels. Based on Russell Group advice that students wishing to keep their options open for admission to the largest number of courses at their member universities should do at least two facilitating subjects, categorical outcome variables are observed for students with grades A* to U in at least two facilitating subjects (the facilitating outcome). The second outcome is whether students chose at least two subjects from the 'more limited suitability' and 'less effective preparation' categories combined (the 'less suitable' outcome). 'More limited suitability' subjects are those where at least one Russell Group member publishes reservations about the subject as appropriate preparation, but on the other hand it is described as useful, recommended or essential for at least one Russell Group course. 'Less effective preparation' subjects are those where reservations are expressed on at least one general list, and no Russell Group courses in related subjects describe the subject as useful, recommended or essential. Only a small number of candidates enter for 'less effective preparation' subjects and all but three of these subjects are in the process of being withdrawn in current A-level reforms (Ofqual, 2014). I therefore combine them with 'more limited suitability' subjects. Figure 1 shows that only a very small proportion of Russell Group entrants hold at least two 'less suitable' subjects, which suggests that they may be unhelpful for entry to Russell Group courses, and so investigating their relationship with student background is valuable. The full list of A-levels in each category is given in Table S1, which also indicates those to be withdrawn.

The outcomes examined are dichotomous, and might generally be modelled using binary logistic regression. To answer the fourth research question, the unobservable effects of schools on subject choices are examined through the use of a fixed effect model. Binary logistic fixed effect models for the full sample do not converge as there are some 2,200 school fixed effects to estimate. I therefore adopt the approach used by Chowdry et al. (2013) who note that where the probability of an outcome is between 0.25 and 0.75, linear probability models give a close approximation to the logit model. In the case of the facilitating subject choice outcome the probability is 0.44. For the second outcome the probability is 0.23. This is slightly outside the suggested limits, but robustness checks of a fixed effect logit model on a smaller sample suggest the same substantive conclusions as the linear probability model. The use of robust standard errors deals with the problem of heteroskedastic residuals. I therefore use linear probability models throughout, so that coefficients are interpreted as the increase in probability

of a student taking at least two facilitating subjects, or at least two more limited suitability/less effective preparation ('less suitable') subjects, for a unit increase in the variable of interest, all else equal. In models 1 to 3 the data are analysed at the disaggregated level. The grouping of students within schools is taken into account by using clustered standard errors.

Model 1 shows the relationship between social background and subject choice without taking any account of the pathways through which the association might be mediated. I then control for individual observable characteristics that may contribute to the association. These are added in four blocks. Model 2a includes confounding demographic variables, likely to be associated both with SES and subject choice, but not forming part of the pathway from SES to choice (ethnicity, gender, region, English as additional language (EAL), special educational needs (SEN) status).

Prior attainment is then taken into account (models 2b-2d), as evidence suggests students with high prior attainment are more likely to choose maths and science subjects (all facilitating subjects) at A-level (Gill & Bell, 2013; Vidal Rodeiro, 2007), and conversely the lower their scores at GCSE the more likely students are to choose newer or vocational subjects (Vidal Rodeiro, 2007), all of which are included in the 'less suitable' outcome. This is consistent with work by Coe, Searle, Barnby, Jones, and Higgins (2008), whose work suggests that subjects in the facilitating category are on average more difficult than the mean, and those I classify as "less suitable" category are largely easier. Gorard and See (2008) suggest that, in the context of science subjects, having high prior attainment is likely to mean students are prepared to take hard subjects, both because they think they will succeed at them and because schools may have threshold GCSE results for these subjects at A-level. Many 'less suitable and some 'useful' subjects are not commonly taught at GCSE making the imposition of threshold scores less likely, and so opening up the subject to students with lower attainment. Given that both attainment at GCSE and the choice of subjects (the English EBacc of core academic subjects that are likely to be necessary preparation for the study of many facilitating A-level subjects) is related to SES (Allen, 2015; Sammons et al., 2014; Vidal Rodeiro, Sutch, & Zanini, 2013) it is likely that one of the ways social class is associated with subject choice, particularly of facilitating subjects, is through differential prior attainment and choice. Attainment controls are added sequentially in models 2b to 2d to gain understanding of the relative importance of attainment at age 11 and 16, and the role of having made particular GCSE choices at age 14.

I next take into account the observable characteristics of schools (model 3). Schools can influence students' A-level subject choices in a variety of ways, most obviously by either providing particular subjects or not. There are large differences between the proportion of schools offering particular subjects by school type: facilitating subjects are offered by higher proportions of selective state (grammar) schools than comprehensives and further education (FE) colleges; the converse is generally true of 'less suitable' subjects; and 6th form colleges tend to offer many subjects across all categories, because of their large A-level cohorts. Differences also exist in some subjects by school gender and size, and also mean school attainment, with larger proportions of higher attaining schools offering broadly more facilitating and fewer 'less suitable' subjects (Gill, 2015a). Vidal Rodeiro (2007) notes that independent

and grammar schools offer fewer of the vocational and newer subjects introduced in the early 2000s with the aim of broadening the A-level curriculum. School type has been shown to have an effect on subject choice, with students from selective maintained schools, independent schools and colleges being more likely to take science A-levels, those in further education and tertiary colleges arts, social sciences and humanities, and from comprehensive schools a mixture (Vidal Rodeiro, 2007). The mechanism by which this happens is not clear and no account was taken of the choice set provided by type of school. I therefore control for a vector of school level variables relating to breadth of subject choice selectivity, school type, size of the A-level cohort and school gender. These school level characteristics may act as pathways mediating the relationship between SES and subject choice, but this has not been examined to date. They are related to background as different types of students are found in different schools.

Model 3 includes only those school level variables that are observed in the data. A review of the literature by Bennett, Braund, and Sharpe (2013) in the context of STEM subject uptake suggests that a range of unobservable school level characteristics may play a part in subject choice. These include school ethos, leadership and management, curriculum effects, A-level entry policies, careers advice and guidance, the availability of enrichment activities and the impact of specialist teachers. These unobserved characteristics are also likely to act as mediating pathways, and are therefore taken into account in the final formulation of the model, model 4, which is defined as follows, using a linear probability model for outcome y_{ijt} :

$$\hat{P}(y_{ijt}=1 | S_{ij}, T_t, D_{ij}, A_{ij}, I_j) = \hat{y}_{ijt}$$

$$\text{Where } y_{ijt} = \alpha + \beta S_{ij} + \gamma T_t + \delta D_{ij} + \rho A_{ij} + \tau_j I_j + u_j + \varepsilon_{ijt}$$

The coefficient β is interpreted as the change in probability of choosing at least two facilitating subjects (or two 'less suitable' subjects) associated with a change in SES category S_{ij} compared with the baseline (least privileged quintile). T_t is a cohort dummy to account for trends in subject choice over the three pooled cohorts. D_{ij} are demographic variables, A_{ij} is prior attainment and I_j are observed school level characteristics. The u_j s are school level residuals and ε_{ijt} s are individual level residuals.

The unobserved school level residuals are dealt with by using school fixed effects in model 4 to control for all differences in schools, both observed and not, so disentangling the relationship of subject choice with SES from that relating to schooling by effectively comparing students of different backgrounds within the same school.

An alternative multilevel method of estimating both individual and school level associations with subject choice is to use a random effect estimator. Clarke, Crawford, Steele, and Vignoles (2013) suggest that the fixed and random effect estimators should give similar results either where there are large numbers of observations per school, or the variation between schools is small compared with that within schools. The random effects estimator is the weighted average of the between group and within group (fixed effect) estimates – in this case the groups are schools. Both conditions hold in this administrative data; there are large numbers of

observations per school and the proportion rho of the total variation in the outcome measures due to school level clustering is small (0.11 for the facilitating subject outcome and 0.12 for the 'less suitable' outcome). The advantage of using the random effects approach is that it allows coefficients on observed school level variables to be estimated, but it requires a more problematic assumption than fixed effects, that u_j and ε_{ijt} are mutually independent and have zero means given the values of the remaining explanatory variables. In practice this means that unobserved school characteristics which make a difference to subject choice outcomes, such as the IAG given to students, must not be correlated with their social background. As they are likely to be correlated in this case the fixed effects method is used, and the random effect results are noted only.

Data

I use individual level administrative Key Stage 5 (KS5) data for three cohorts from the NPD. This contains detailed attainment data by subject for those students taking one substantial level 3 qualification (defined as at least the size of one A level - 180 guided learning hours per year) in 09/10, 10/11 and 11/12, and individual level characteristics from the spring School Census for students in state schools. These cohorts are linked to KS2 (age 11) attainment data relating to normal progression through school, and to KS4 (GCSE age 16) data with both attainment and School Census variables which are used when School Census data at KS5 is missing. The databases contain school identifiers allowing schools to be matched to Edubase, which provides school level variables (school gender and selectivity).

Table 1: Socio-economic gradient of A-level cohort and KS4 cohorts

All state school students	% in SES quintile 1	% in SES quintile 2	% in SES quintile 3	% in SES quintile 4	% in SES quintile 5	% missing SES	Number of students
KS4 state school cohorts 2007/8 – 2009/10	19.2	19.2	19.2	19.2	19.1	4.1	1,803,337
KS5 at least one A-level 2009/10 – 2011/12	8.8	13.3	18.5	23.9	32.1	3.4	623,575
KS5 at least three A-levels 2009/10 – 2011/12	7.5	12.1	17.8	24.3	35.0	3.4	485,252
Common estimation sample used in analysis	7.6	12.4	18.5	25.4	36.1	0	444,467

The sample is restricted to those taking three or more A-levels excluding general studies and critical thinking, as they are those with a realistic chance of admission to a highly selective university. I recognise that by only including those doing A-levels a large proportion of lower SES students are excluded from analysis because of prior selection, but my aim is specifically to examine choices made by those who have already decided to stay in the post-compulsory academic cohort and who are most likely to attend university. Table 1 shows the socio-economic gradient of state school students taking A-levels compared with the year 11 state

school cohort as a whole: those staying on to take at least one A-level are privileged compared with their year 11 peers with only 8.8% coming from the bottom year 11 quintile and 32.1% coming from the top, and the three A-level cohort is slightly more privileged still.

To be in a cohort, students must be completing their studies in that year, and be aged 18 or younger on 31 August. Those taking the International Baccalaureate or Cambridge Pre-U qualification without three A-levels are excluded from this analysis because of their small numbers (Vidal Rodeiro et al., 2013). Students taking BTEC are also excluded as they represent a very small proportion of English 18-year-old entrants to high tariff British universities; 2% of those accepted in 2014 hold the equivalent of ABB or better from BTEC compared with 77% from A-levels (UCAS, 2014).

If a student does four A-levels it is possible that they appear as positives for both outcomes. Only 1,637 students in the sample do at least two facilitating and at least two 'less suitable' subjects. The analysis of the at least two 'less suitable' subject' outcome was rerun, excluding these students. For all model specifications SES coefficients were within a very small margin (.001) of those for the full sample.

Table 2: A-level subject choices by student characteristics

Estimation sample	% taking at least two facilitating subjects	% taking at least two 'less suitable' subjects
OVERALL	44.4	23.1
BY GENDER		
Female	38.8	23.6
Male	51.5	22.5
BY FREE SCHOOL MEAL ELIGIBILITY:		
In either year 13 or year 11	37.3	25.6
In neither year	44.7	23.0
BY SES QUINTILE		
Quintile 1	37.1	26.3
Quintile 2	41.2	25.0
Quintile 3	44.1	23.6
Quintile 4	47.3	21.9
Quintile 5	52.0	18.9
BY ETHNICITY		
Any other ethnic group	55.5	18.4
Asian	52.7	18.2
Black	36.0	22.8
Chinese	64.0	17.4
Mixed	44.7	21.7
Undeclared	46.2	21.7
White	43.5	23.8

Table 2 shows differences in A-level outcome by student characteristics. Male students are much more likely to take facilitating subjects (51.5% compared with 38.8% of females), although only account for 43.7% of those taking three A-levels. Chinese students are the ethnic group much the most likely to take facilitating subjects (64.0%) followed by Asian students (52.7%) with white students (43.5%) and black students (36.0%) less likely. Much smaller differences in the choice of 'less suitable' subjects are observed by gender, but ethnicity remains important; Chinese students are the least likely to choose these subjects, and white students the most.

The individual data on SES available in the administrative datasets is limited, with the only individual level variable being free school meals eligibility (FSM). Hobbs and Vignoles (2010) suggest that FSM status is a poor proxy for income, and it only allows comparison of those who

are at the bottom of the SES distribution (in the case of those doing three A-levels, 5.3%) with those above. The raw difference in choosing facilitating subjects by this measure is 7.4pp, and 2.6pp in 'less suitable' subjects.

In order to investigate the SES gradient across the whole distribution I follow Chowdry et al. (2013) in using a combination of individual FSM data with neighbourhood data to construct a measure of SES. I note that this has problems associated with it (for example being on FSM indicates a level of income deprivation which is already likely to be taken into account in some of the neighbourhood variables (Chowdry et al., 2013)). As a proxy it is a noisy measure with the error term in its measurement creating an endogeneity bias in the regression equation, skewing the estimated coefficients towards zero (attenuation bias). The models are therefore likely to show a smaller relationship in absolute terms between the outcomes of interest and SES than is actually the case. I use principal components analysis to construct an index of socio-economic background combining the following measures, linked using the student's home postcode at age 16 where available, and where missing at age 18. The variables used to construct the index are:

- Whether a student is eligible for FSM at either or both of age 16 and 18;
- An index of multiple deprivation (available for neighbourhoods containing around 700 households);
- The classification of residential neighbourhoods type, based on individual postcodes, and derived from information on housing details and socio-economic characteristics (each postcode contains around 15 households)
- Local area measures for around 150 households based on the 2011 census, of the proportion of:
 - individuals working in higher or lower professional or managerial occupations;
 - individuals aged 16 and over whose highest educational qualification is national qualification framework level 3 (ie A-level or equivalent) or above;
 - households that own their home.

Although it is problematic to use dichotomous variables in a principal component analysis (Kolenikov & Angeles, 2009), only one of the variables in this analysis is binary, and the remainder are continuous, so reducing the problem. Chowdry et al. (2013) construct an SES measure using this method and find their results substantively unchanged if they use FSM together with each of the measures separately. The Kaiser-Meyer-Olkin measure of sampling adequacy for the index I derive using this method is 0.8, suggesting that it is appropriate to consider the common variance as a measure of a latent underlying variable, in this case SES. The three A-level cohort of state school students is split into quintiles according to this measure, with a mean value of 3 across cohorts.

I observe a range of school level characteristics from Edubase. I classify schools as FE and 6th form colleges and selective or non-selective mainstream state schools. Schools are matched to individual students through their combined Local Authority and Establishment numbers, which remain constant when a school's status changes.

Table 3: Descriptive statistics by school type

Estimation sample	Non selective state schools	Selective state schools	6 th form colleges	FE colleges	Overall
Percentage taking at least two facilitating A levels	45.7	65.5	35.3	29.5	44.4
Percentage taking at least two 'less suitable' A-levels	23.1	9.3	28.2	29.4	23.1
Mean SES quintile (in sample)	3.06	3.27	2.85	2.62	3.00
Mean SES quintile (KS4 cohort)	3.77	3.96	3.53	3.34	3.70
Mean capped GCSE points, including equivalents	396	423	397	392	399
Mean 'facilitating' GCSE passes A*-C per student	4.3	5.8	4.1	3.9	4.4
Mean 'less suitable' GCSE passes A*-C per student	1.7	1.6	1.7	1.7	1.7
Mean total GCSE passes A*-C per student	8.2	9.4	8.1	7.8	8.3
Mean number of facilitating A-level subjects offered by school	7.7	11.0	11.2	6.3	8.0
Mean number of 'less suitable' A-level subjects offered by school	8.2	7.2	18.3	8.4	8.6
Mean total number of A-level subjects offered	20.6	25.5	40.7	20.4	21.8
Number of students	250,053	53,954	103,631	36,829	444,467
Number of schools/colleges	1,752	162	95	167	2,176

Table 3 shows the large differences in subject choice by school type, for those taking at least three A-levels. More than twice the proportion of students at selective schools take at least two facilitating subjects than those at FE colleges, but there are also substantial differences between non selective state schools and 6th form and FE colleges, despite the relatively similar average GCSE attainment and GCSE subject choice pattern at these types of schools. The converse is observed for 'less suitable' A-levels, with students at colleges around three times as likely to take them than those at selective state schools, and over 6pp more likely than those at non selective schools.

The number of subjects from different categories offered varies considerably by school type. 6th form colleges typically offer large numbers of A-levels, so facilitating subjects form a relatively small proportion of average provision (28%), compared with 43% at selective schools. Conversely, the mean proportion of 'less suitable' subjects offered by selective schools is 28%, compared with 40% or more at non-selective schools and colleges. The SES quintile measures given in Table 3 show that students at colleges are on average less privileged than those at non-selective schools, who in turn are less privileged than those at selective schools.

Two measures are used for school choice set in each set of models; the total number of A-level subjects offered in the three-year period, and the number of subjects offered from the outcome of interest. A subject is counted within a school's offer if at least three students have taken it during the three years. A limitation of these measures is that all A-level subjects provided by a school are included, where in practice there may be timetabling or other constraints on choice. Evidence from the Longitudinal Study of Young People in England 2005-6 suggests that around a fifth of KS4 students are not able to take preferred subjects at GCSE; in the majority of cases because of timetable clashes or full classes (Jin, Muriel, & Sibiet, 2011). A similar effect at A-level has not yet been investigated, but it is possible that the 'offer' measures overstate the choices actually available to students.

Although A-level subject choices are often made at the end of year 11, I use the school characteristics of the 6th form attended because those changing school for A-levels are likely to make decisions based on available subjects at the school or college they plan to attend.

A rich set of data is available with which to control for prior attainment. KS2 deciles are constructed according to average point scores in English, maths and science. The relationship of subject choice with KS2 deciles is linear apart from for the top decile. I therefore include standardized average points scores English, maths and science, and a dummy to indicate top decile in KS2. KS1 quintiles are not included in the models as they have negligible effect on model fit, yield largely non-significant coefficients and reduce the common estimation sample by some 20,000 relating to A-level students who could not be found in the KS1 data 11 years earlier and had no other missing variables. Standardized capped GCSE and equivalents points scores are used as overall KS4 controls. An indicator of the number of facilitating A-level subjects a student could potentially have taken is constructed using the number of related 'facilitating' GCSE subjects held at grades A*-C (maths, English literature, separate sciences, languages, history and geography). The number of GCSE grades A*-C in subjects corresponding to 'less suitable' A-level subjects in the taxonomy is also counted to see whether having done these subjects from 14-16 is important in choice of A-levels.

Models are run on 444,467 complete cases from 485,252 students with at least three 'counting' A-levels over the three cohorts. This loss of data and therefore statistical power is not a significant problem given the size of the administrative dataset. Listwise deletion generally results in estimated standard errors that are good estimates of the true ones (Allison, 2001). The missing cases are slightly more likely to choose at least two facilitating subjects, and less likely to choose 'less suitable' ones, whilst being slightly less privileged. Their absence from the analysis would therefore, if anything, increase the SES gradients observed, so tending to

overstate the relationship between subject choice and social background. The missing data is of three main kinds; first, missing data from the School Census (LSOA, ethnicity, FSM, SEN status for some 12,000 students, disproportionately at FE colleges so likely to be lower SES, on average, second, 3,500 students appearing in KS5 not matched to KS4, and finally around 24,000 students with missing attainment data at KS2.

Results

Table 4: Gradient in choice of at least two facilitating A-levels by SES

At least two facilitating A Levels	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 3	Model 4
SES quintile 2	0.041*** (0.004)	0.047*** (0.003)	0.025*** (0.003)	0.009*** (0.003)	0.000 (0.002)	-0.002 (0.002)	-0.003 (0.002)
SES quintile 3	0.070*** (0.005)	0.079*** (0.004)	0.047*** (0.003)	0.017*** (0.003)	0.002 (0.003)	-0.001 (0.003)	0.000 (0.002)
SES quintile 4	0.102*** (0.005)	0.114*** (0.004)	0.072*** (0.003)	0.029*** (0.003)	0.009** (0.003)	0.004 (0.003)	0.005* (0.002)
SES quintile 5	0.149*** (0.007)	0.160*** (0.006)	0.103*** (0.004)	0.044*** (0.004)	0.015*** (0.004)	0.007* (0.003)	0.011*** (0.003)
Standardized maths score age 11			0.102*** (0.001)	0.040*** (0.001)	0.031*** (0.001)	0.032*** (0.001)	0.032*** (0.001)
Standardized English score age 11			0.000 (0.001)	-0.035*** (0.001)	-0.040*** (0.001)	-0.040*** (0.001)	-0.040*** (0.001)
Standardized science score age 11			0.079*** (0.001)	0.042*** (0.001)	0.031*** (0.001)	0.031*** (0.001)	0.030*** (0.001)
Top decile age 11			0.121*** (0.003)	0.049*** (0.003)	0.047*** (0.003)	0.046*** (0.003)	0.045*** (0.003)
Standardized GCSE capped points				0.215*** (0.002)	0.177*** (0.002)	0.177*** (0.002)	0.177*** (0.002)
Total number GCSEs A*-C					-0.023*** (0.001)	-0.023*** (0.001)	-0.026*** (0.001)
Number of 'facilitating' GCSEs A*-C					0.079*** (0.001)	0.079*** (0.001)	0.085*** (0.001)
Number of facilitating A-levels offered						0.012*** (0.002)	
Total A-level subjects offered						-0.005*** (0.000)	
School gender – boys (baseline mixed)						-0.037*** (0.009)	
School gender – girls (baseline mixed)						0.018* (0.008)	
School type – selective school (baseline non-selective school)						-0.065*** (0.008)	
School type – 6 th form college (baseline non-selective school)						-0.048*** (0.008)	
School type – FE college (baseline non-selective school)						-0.079*** (0.008)	
School size (per 100 students)						0.001 (0.002)	
R ²	.0112	.0423	.1681	.2745	.3177	.3263	.3140
% with predicted values <0 or >1						8	8
Number of schools							2,176
Cohort controls	X	X	X	X	X	X	X
Demographic controls		X	X	X	X	X	X
School fixed effects							X
Observations	444,467	444,467	444,467	444,467	444,467	444,467	444,467

Table 5: Gradient in choice of at least two 'less suitable' A-levels by SES

At least two 'less suitable' A Levels	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 3	Model 4
SES quintile 2	-0.013*** (0.003)	-0.021*** (0.003)	-0.005 (0.003)	0.004 (0.002)	0.008** (0.002)	0.010*** (0.002)	0.009*** (0.002)
SES quintile 3	-0.028*** (0.004)	-0.040*** (0.003)	-0.017*** (0.003)	-0.001 (0.003)	0.006* (0.003)	0.011*** (0.003)	0.010*** (0.002)
SES quintile 4	-0.045*** (0.004)	-0.061*** (0.004)	-0.030*** (0.003)	-0.008** (0.003)	0.003 (0.003)	0.009*** (0.003)	0.009*** (0.002)
SES quintile 5	-0.075*** (0.005)	-0.092*** (0.004)	-0.053*** (0.004)	-0.021*** (0.004)	-0.005 (0.003)	0.005 (0.003)	0.006** (0.002)
Standardized maths score age 11			-0.044*** (0.001)	-0.012*** (0.001)	-0.004** (0.001)	-0.004** (0.001)	-0.003* (0.001)
Standardized English score age 11			-0.045*** (0.001)	-0.026*** (0.001)	-0.018*** (0.001)	-0.018*** (0.001)	-0.017*** (0.001)
Standardized science score age 11			-0.058*** (0.001)	-0.038*** (0.001)	-0.029*** (0.001)	-0.028*** (0.001)	-0.027*** (0.001)
Top decile age 11			-0.003 (0.003)	0.035*** (0.002)	0.032*** (0.002)	0.034*** (0.002)	0.031*** (0.002)
Standardized GCSE capped points score				-0.110*** (0.002)	-0.095*** (0.002)	-0.072*** (0.002)	-0.072*** (0.001)
Total number GCSEs A*-C					-0.048** (0.001)	-0.046*** (0.001)	-0.050*** (0.001)
Number of 'less suitable' GCSEs A*-C					0.096*** (0.001)	0.093*** (0.001)	0.098*** (0.001)
Number of 'less suitable' A-levels offered						0.018*** (0.001)	
Total A-level subjects offered						-0.008*** (0.001)	
School gender – boys (baseline mixed)						0.025*** (0.007)	
School gender – girls (baseline mixed)						0.009 (0.006)	
School type – selective school (baseline non-selective school)						0.050*** (0.007)	
School type – 6 th form college (baseline non-selective school)						0.021* (0.009)	
School type – FE college (baseline non-selective school)						0.022** (0.008)	
School size (per 100 students)						0.003 (0.002)	
R ²	.0049	.0122	.0936	.1347	.1903	.1996	.1877
% with predicted values <0 or >1						11	10
Number of schools							2,176
Cohort controls	X	X	X	X	X	X	X
Demographic controls		X	X	X	X	X	X
School fixed effects							X
Observations	444,467	444,467	444,467	444,467	444,467	444,467	444,467

SES and choice of subjects

Model 1 in Tables 4 and 5 gives the marginal effects from the baseline models, showing the raw results of each of both outcomes, conditioning only on cohort. The socio-economic gap is clear; students from the top SES quintile are 14.9pp more likely than those in the bottom quintile to take at least two facilitating subjects and 7.5pp less likely to take at least two 'less suitable' subjects. These are large differences, given that the overall probability of taking two facilitating subjects is 44.4% and 'less suitable' 23.1pp. A clear gradient in outcomes across SES quintiles can be seen.

Models 2a to 2d of each table show the effect of adding individual level controls, first demographic and then attainment. The addition of demographic controls (model 2a) slightly accentuates the gradient for facilitating subjects to 16.0pp between the top and bottom SES quintiles, dealing with the confounding caused by heterogeneity in demographic covariates by SES. The gradient in choice of 'less suitable' subjects is also increased, to -9.2pp. There are noteworthy differences in subject choice by gender and ethnicity for both outcomes, which will be the subject of future study.

SES, attainment and choice of subjects

Model 2b of each table shows the results of conditioning on attainment aged 11. A substantial proportion of the observed raw difference in facilitating subject choice (Table 4) is mediated by attainment; controlling just for KS2 attainment reduces the SES gap by 5.7pp. Adding GCSE (and equivalents) scores in model 2c reduces the gap by another 5.9pp, which is consistent with facilitating subjects being considered hard. KS2 scores become less important, but an increase of one standard deviation in GCSE scores (37 points with mean 399) is associated with a 21.5pp increase in probability of taking two facilitating A-levels. Model 2d demonstrates the importance of taking GCSE subjects that provide suitable preparation for facilitating A-levels. Controlling for total number of GCSEs, having just one more 'facilitating' GCSE rather than any other is associated with a 7.9pp increase in chance of taking two facilitating A-levels. The role of the overall GCSE score is still important, but less than before. For a student of given GCSE overall attainment and 'facilitating' GCSEs, having extra (useful or 'less suitable') GCSEs is negatively associated with choosing two facilitating A-levels. Controlling for attainment and subject choice at GCSE almost entirely accounts for the gap in A-level facilitating subject choice by SES. Only the top two quintiles differ from the bottom at conventional significance levels, and the gap between the top and bottom quintiles is very small. It is possible that there is some endogeneity bias here; students wishing to take facilitating A-levels for which some hurdle mark has been set may work harder to improve their GCSE score, or choose 'facilitating' GCSE subjects at 14+, and such foresight may be related to social background. If so, the role of GCSEs may be slightly overstated and that of SES understated, but the overall relationship of subject choice with GCSEs would remain much more important than with SES.

Prior attainment is also seen to have a role in choice of at least two 'less suitable' subjects in Table 5, consistent with their being easier subjects on average. Higher KS2 scores are associated with being less likely to make this choice. Controlling for KS2 results reduces the

negative association of higher SES quintile with these subjects, because of the relationship of social background with attainment. Overall GCSE scores are negatively associated with the choice of two 'less suitable' subjects, and model 2c suggests they are more important than KS2 scores. As might be expected, model 2d shows that taking subjects at GCSE corresponding to A-level 'less suitable' subjects makes it more likely that such subjects are taken at A-level too. For students of given overall GCSE score and number of 'less suitable' GCSEs, having more (facilitating or useful) GCSEs is associated with a lower probability of choosing two 'less suitable' A-levels. Including attainment controls completely accounts for the difference between uptake of two 'less suitable' A-levels by social background.

SES, schools and choice of subjects

Model 3 of Tables 4 and 5 includes the addition of school level observable variables. The inclusion of these variables removes the small remaining SES gradient choice of facilitating subjects, indicating some small further mediation by schools of the relationship between social background and subject choice, controlling for ability and demographics.

Whilst Model 3 cannot account for unobserved school effects, it has the benefit of showing how school level observable characteristics in the model relate to subject choice. The first significant school level predictor in each table is the number of subjects by category offered by the school. Tables 4 and 5 suggest the choice of two facilitating two 'less suitable' subjects is sensitive to the number offered, with an increase in uptake of 1.2pp and 1.8pp respectively per additional subject from the category offered, keeping the total number of A-level subjects offered fixed. So, for example, offering five more 'less suitable' subjects rather than useful or facilitating ones, is associated with around 9pp more students taking at least two 'less suitable' ones, from an average of 23.1pp. Holding the number of facilitating subjects constant, increasing the total number of subjects offered is associated with a smaller but significant decrease in probability of taking at least two facilitating subjects (-0.5pp) and a similar but slightly larger relationship is seen with 'less suitable' subjects (-0.8pp).

Students at all boys' schools are 3.7pp less likely to take at least two facilitating subjects (Table 4) than those at mixed schools, and conversely 2.5pp more likely to take at least two 'less suitable' ones, controlling for all else.

A significant predictor for both outcomes is school type; whether a student attends a selective or non-selective school, FE or 6th form college. Students at FE colleges, 6th form colleges and selective schools are less likely than those at a non-selective school to take at least two facilitating subjects (-7.9pp, -4.8pp and -6.5pp respectively), after controlling for attainment and breadth of choice. Students at selective schools, 6th form colleges and FE colleges are all more likely to take at least two 'less suitable' subjects. The raw differences observed in subject choice by school type in Table 3 are explained by the difference in attainment, GCSE subject choice, and breadth of A-level offer.

Controlling for unobserved school variables through the fixed effect model 4 shows substantially the same story as model 3 for both outcomes. The SES gap remains very small, with only a 1.1pp difference between the top and bottom quintile for facilitating subjects, and an insignificant gap (at 0.1% confidence) between top and bottom for 'less suitable' A-levels. Unobserved

characteristics of schools do not seem to be important in accounting for the SES gap once attainment and GCSE subject choice are taken into account.

Results from the random effects version of model 4 give very similar results to the fixed effect model for the SES gradient for both outcomes.

Discussion and conclusion

This research contributes to the literature by using a new taxonomy to illustrate a clear difference in A-level subject choice patterns by social background. Students from less privileged backgrounds are less likely to choose those subjects described as particularly helpful for highly selective university entry and more likely to choose those that are not, so potentially limiting their future educational trajectory. But these raw differences in A-level choice by social background are effectively removed once attainment and the schools and colleges attended are taken into account, with GCSE subject choices and performance seen to be particularly important.

Although a detailed examination of the usefulness of A-levels for university entry from the categories developed in my taxonomy is the subject of a future paper, *prima facie* there is a relationship between holding A-levels from particular categories and entry to a highly selective university for students from these cohorts. The differentials in subject choice observed across SES quintile in this study may therefore form part of the reason for the under-representation of students from less privileged backgrounds at highly selective universities.

The models show that even when attainment at age 11 is taken into account, progress made between KS2 and KS4 is significant in the choice of A-level subjects, and it is well established that such progress is related to social background, with children from less privileged backgrounds making less progress on average than their more privileged peers (Allen, 2015; Sammons et al., 2014), even when they have been high attaining at age 11 (Crawford et al., 2015). Having good GCSE results is associated with higher chances of taking at least two facilitating subjects, which is consistent with these subjects being considered more difficult than others. Differential take-up is then likely both through schools setting hurdle marks and through students' own sense of whether they are likely to succeed. But it is not just general GCSE attainment as a measure of 'capacity to learn' that matters: particular GCSE subject choice is important too, and this work suggests decisions made at age 14 may have a lasting impact on individual's life chances. Here, too, we know there are differentials by social background, with students from poorer backgrounds less likely to choose the subjects that will provide good preparation for taking facilitating A-level subjects (Allen, 2015), suggesting an important role for information, advice and guidance at age 14. Providing A-level subject choice guidance at age 16, such as 'Informed Choices', may be too late. The GCSE subject choice problem is being addressed to an extent through the introduction of the EBacc, but it will take some time for this to work through: the first cohort for which the EBacc is compulsory will take GCSEs in 2020.

The main way schools can mediate the relationship between social background and A-level subject choice is through reducing the GCSE attainment and age 14 subject choice gap. This work suggests that other characteristics of schools are relatively unimportant in closing the SES

gap, although they can make a difference to A-level subject choices. Most obviously, in order to meet the needs and interests of their particular students and because of the size of their cohort, schools will make different decisions about the subjects they offer. Providing choice is desirable, but can also result in decisions which limit future options. It therefore seems particularly important that where students have a large choice of subjects, good IAG is available to help them decide which to take. The significant negative association of taking two facilitating A-levels with going to FE or 6th form college, even after taking account of the number of such subjects offered, suggests that particular efforts might be needed to make sure that students going to colleges are making well informed decisions. This may be challenging given that many are likely to be going to college specifically to take subjects not available at their existing school, and that the opportunities for the provision of IAG may be limited before students arrive to take up their studies. Grammar schools are an interesting case: the high proportions of students taking at least two facilitating A-levels (and conversely not taking 'less suitable' A-levels) is because of the high GCSE scores and 'facilitating' GCSE choices of their students and the weight in A-level provision towards facilitating subjects. Net of these relationships, students at grammar schools are less likely to take at least two facilitating subjects and more likely 'less suitable' than non-selective school students.

The cohorts in this study chose their A-level subjects before the Government's AAB in at least two facilitating subjects performance indicator was introduced. Whether this indicator makes a difference first to the proportion of state school students taking facilitating A-levels, and second to the proportion of state school entrants to highly selective universities remains to be seen. The measure is likely to cause shifts away from useful, more limited suitability and less effective preparation subjects and towards facilitating ones, at least for high attaining students, and indeed almost all 'less effective preparation' subjects are being removed from teaching under current reforms (Ofqual, 2014). But the introduction of this measure is unlikely to solve the more fundamental problem with A-level subject choice suggested by this paper, that differential attainment and GCSE choice by social background by students equally well qualified at age 11 contributes significantly to the observed differences in A-level choice. The evidence suggests that students staying on to take three A-levels will choose facilitating subjects, from whatever background they come, if they have good GCSE scores in helpful subjects. Improving both the overall attainment of lower SES students and ensuring they have good advice on GCSE subject choice at age 14+, as well as good advice on A-level subject choice at 16 (particularly where there are a large number of options open to them) seem the most likely routes to address the subject choice gap.

Further work is needed to examine the extent to which A-level subject choice accounts for the gap in admission to highly selective university by SES, and the extent to which this varies by course applied for, as well as whether the introduction of the facilitating subject performance indicator achieves its intended aims. Examining further the unexplained school level differences, for example the availability of specialist A-level teachers through linking this data with the School Workforce Census, as well as individual level differences in subject choice such as ethnicity also seem fruitful areas for further study.

References

- Allen, R. (2015). *Missing Talent*. (London, Sutton Trust). Available online at: <http://www.suttontrust.com/wp-content/uploads/2015/06/Missing-Talent-final-june.pdf> (accessed 29 April 2016.)
- Allison, P. D. (2001) Missing data. *Sage university papers series on quantitative applications in the social sciences* (Thousand Oaks, CA, Sage).
- Anders, J. (2012) The link between household income, university applications and university attendance, *Fiscal Studies*, 33(2), 185-210.
- Becker, G. S., & Tomes, N. (1986) Human Capital and the Rise and Fall of Families, *Journal of labor economics*, 4(3), 1.
- Bennett, J., Braund, M., & Sharpe, R. (2013) *Student attitudes, engagement and participation in STEM subjects* (London, The Royal Society/ University of York). Available online at: [https://pure.york.ac.uk/portal/en/publications/student-attitudes-engagement-and-participation-in-stem-subjects\(0034c02e-86e6-4fc6-9eed-d7d24c7a0e79\).html](https://pure.york.ac.uk/portal/en/publications/student-attitudes-engagement-and-participation-in-stem-subjects(0034c02e-86e6-4fc6-9eed-d7d24c7a0e79).html) (accessed January 5 2016).
- Blanden, J., Gregg, P., & Macmillan, L. (2007) Accounting for intergenerational income persistence: noncognitive skills, ability and education, *The Economic Journal*, 117(519), C43-C60.
- Blanden, J., & Macmillan, L. (2014) *Education and intergenerational mobility: help or hindrance?* (London, Centre for Analysis of Social Exclusion, London School of Economics). Available online at: http://eprints.lse.ac.uk/58045/1/spccwp08_full%20text.pdf (accessed 28 March 2015).
- Boliver, V. (2013) How fair is access to more prestigious UK universities?, *The British Journal of Sociology*, 64(2), 344-364.
- Britton, J., Dearden, L., Shephard, N., & Vignoles, A. (2016). *How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background*. (London, Institute for Fiscal Studies). Available online at: <http://www.ifs.org.uk/uploads/publications/wps/wp1606.pdf> (accessed 29 April 2016).
- Burgess, S., & Briggs, A. (2010) School assignment, school choice and social mobility, *Economics of Education Review*, 29(4), 639-649.
- Carneiro, P., & Heckman, J. J. (2002) The evidence on credit constraints in post-secondary schooling, *The Economic Journal*, 112(482), 705-734.
- Cheng, Y., Payne, J., & Witherspoon, S. (1995) *Science and Mathematics in full-time education after 16* (London, Department for Education and Employment).
- Chevalier, A., & Conlon, G. (2003) *Does it pay to attend a prestigious university?* (London, Centre for the Economics of Education, London School of Economics). Available online at: <http://cee.lse.ac.uk/ceedps/ceedp33.pdf> (accessed 6 June 2014).
- Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013) Widening participation in higher education: analysis using linked administrative data, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 431-457.

- Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2013) Revisiting fixed-and random-effects models: some considerations for policy-relevant education research, *Education Economics*, 23(3), 259-277.
- Coe, R., Searle, J., Barmby, P., Jones, K., & Higgins, S. (2008) *Relative difficulty of examinations in different subjects* (Durham, Centre for Evaluation and Monitoring, Durham University).
- Crawford, C. (2012) *Socio-economic gaps in HE participation: how have they changed over time?* (London, Insitute for Fiscal Studies). Available online at: <http://www.ifs.org.uk/bns/bn133.pdf> (accessed 29 September 2014).
- Crawford, C., Johnson, P., Machin, P., & Vignoles, A. (2011) *Social mobility: a literature review* (London, Department for Business, Innovation and Skills). Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32111/11-750-social-mobility-literature-review.pdf (accessed 10 October 2014).
- Crawford, C., Macmillan, L., & Vignoles, A. (2015) *When and why do initially high attaining poor children fall behind?* (London, Centre for Analysis of Social Exclusion, London School of Economics). Available online at: <http://sticerd.lse.ac.uk/dps/case/spcc/WP20.pdf> (accessed 8 May 2016).
- Davies, P., Qiu, T., & Davies, N. M. (2014) Cultural and human capital, information and higher education choices, *Journal of Education Policy*, 29(6), 804-825.
- Davies, P., Telhaj, S., Hutton, D., Adnett, N., & Coe, R. (2008). Socioeconomic background, gender and subject choice in secondary schooling, *Educational Research*, 50(3), 235-248.
- Department for Education (2014) *Facilitating subjects for the 'AAB' 2014 key stage 5 performance table indicator* (London, Department for Education). Available online at: http://www.education.gov.uk/schools/performance/2012/16to18_12/KS5Facil.pdf (accessed 2 March 2015).
- Dilnot, C. (2015). *A Taxonomy of A-Level Subjects According to the Expressed Preferences of Russell Group Universities: Who Does What?* (London, UCL Institute of Education, Department of Quantitative Social Science). Available online at: <http://repec.ioe.ac.uk/REPEc/pdf/qsswp1512.pdf> (accessed 6 January 2016).
- Gill, T. (2015a). *Provision of GCE A level subjects 2014*. (Cambridge, Cambridge Assessment). Available online at: <http://www.cambridgeassessment.org.uk/Images/222146-provision-of-gce-a-level-subjects-2014.pdf> (accessed 24 April 2016).
- Gill, T. (2015b). *Uptake of GCE A level subjects 2014*. (Cambridge, Cambridge Assessment). Available online at: <http://www.cambridgeassessment.org.uk/Images/222143-uptake-of-gcse-subjects-2014.pdf> (accessed 24 April 2016).
- Gill, T., & Bell, J. F. (2013) What Factors Determine the Uptake of A-level Physics?, *International Journal of Science Education*, 35(5), 753-772.
- Gorard, S., & See, B. H. (2008) Is science a middle-class phenomenon? The SES determinants of 16–19 participation, *Research in Post-Compulsory Education*, 13(2), 217-226.
- Gorard, S., & See, B. H. (2009) The impact of socio-economic status on participation and attainment in science, *Studies in Science Education*, 45(1), 93-129.
- Gorard, S., See, B. H., & Smith, E. (2008) The impact of SES on participation and attainment in science - an analysis of available data, in: Royal Society (Ed) *Exploring the Relationship between*

- socioeconomic status and participation and attainment in science education* (London, Royal Society).
- Hobbs, G., & Vignoles, A. (2010) Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36(4), 673-690.
- Hussain, I., McNally, S., & Telhaj, S. (2009) *University quality and graduate wages in the UK* (London, Centre for the Economics of Education, London School of Economics). Available online at: <http://cee.lse.ac.uk/ceedps/ceedp99.pdf> (accessed 6 June 2014).
- Iannelli, C. (2013) The role of the school curriculum in social mobility, *British Journal of Sociology of Education*, 34(5-6), 907-928.
- Jackson, M., & Jonsson, J. (2013) Why does inequality of educational opportunity vary across countries? Primary and secondary effects in comparative context, in: M. Jackson (Ed) *Determined to Succeed? Performance Versus Choice in Educational Attainment* (Stanford, Stanford University Press).
- Jin, W., Muriel, A., & Sibieta, L. (2011). *Subject and course choices at ages 14 and 16 amongst young people in England: insights from behavioural economics*. (London, Department for Education). Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182677/DFE-RR160.pdf (accessed 29 April 2016).
- Kolenikov, S., & 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.
- Marcenaro-Gutierrez, O., Galindo-Rueda, F., & Vignoles, A. (2007) Who actually goes to university?, *Empirical Economics*, 32, 333-357.
- Office of the Deputy Prime Minister (2013) *Social Mobility Indicators*. Available online at: <https://www.gov.uk/government/publications/social-mobility-indicators> (accessed 2 March 2015)
- Ofqual (2014) *An update on the reforms being made to AS qualifications and A levels* (Coventry, Office of Qualifications and Examinations Regulation). Available online at: <https://qualifications.pearson.com/content/dam/pdf/A%20Level/Reforms/2014-04-08-an-update-on-the-reforms-being-made-to-as-qualifications-and-a-levels.pdf> (accessed 5 August 2015).
- Russell Group (2011) *Informed choices* (London, The Russell Group).
- Russell Group (2012) *Russell Group film advises prospective students on A-level choices* (London, The Russell Group). Available online at: <http://russellgroup.ac.uk/news/russell-group-film-advises-prospective-students-on-a-level-choices/> (accessed 7 March 2015).
- Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). *The Effective Pre-school, Primary and Secondary Education Project (EPPSE 3-16+) Influences on students' GCSE attainment and progress at age 16*. (London, Department for Education). Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/351495/RB352_-_Influences_on_Students_GCSE_Attainment_and_Progress_at_Age_16_Brief.pdf (accessed 18 April 2016).
- Sharp, C., Hutchison, D., Davis, C., & Keys, W. (1996) *The take-up of advanced mathematics and science courses: summary report* (Slough, National Foundation for Educational Research).

- Sullivan, A., Parsons, S., Wiggins, R., Heath, A., & Green, F. (2014) Social origins, school type and higher education destination,. *Oxford Review of Education*, 40(6), 739-763.
- Toth, K., Sammons, P., & Sylva, K. (2015) *Subject to background* (London, Sutton Trust). Available online at: <http://www.suttontrust.com/wp-content/uploads/2015/03/Subject-to-background1.pdf> (accessed 8 April 2016).
- UCAS (2014) *End of cycle report 2014* (Cheltenham, Universities and Colleges Admissions Service). Available online at: <https://www.ucas.com/sites/default/files/2014-end-of-cycle-report-dec-14.pdf> (accessed 25 January 2016).
- Van de Werfhorst, H. G. (2002) A detailed examination of the role of education in intergenerational social-class mobility, *Social Science Information*, 41(3), 407-438.
- Vidal Rodeiro, C. L. (2007) *A level subject choice in England: Patterns of uptake and factors affecting subject preferences*. (Cambridge, Cambridge Assessment). Available online at: <http://www.cambridgeassessment.org.uk/images/111069-a-level-subject-choice-in-england-patterns-of-uptake-and-factors-affecting-subject-preferences.pdf> (accessed 5 February 2015).
- Vidal Rodeiro, C., Sutch, T., & Zanini, N. (2013). *Prior learning of undergraduates in UK higher education institutions*. (Cambridge, Cambridge Assessment). Available online at: <http://www.cambridgeassessment.org.uk/Images/230030-prior-learning-of-undergraduates-in-uk-higher-education-institutions.pdf> (accessed 11 April 2016).
- Walker, I., & Zhu, Y. (2011) Differences by degree: evidence of the net financial rates of return to undergraduate study for England and Wales, *Economics of Education Review*, 30(6), 1177-1186.
- Whitty, G., Hayton, A., & Tang, S. (2015) Who you know, what you know and knowing the ropes: a review of evidence about access to higher education institutions in England, *Review of Education*, 3(1), 27-67.

Supplementary information linked to the online version of the paper at Wiley-Blackwell

Table S1: Taxonomy of A-levels available for teaching in 2014/15

<i>Facilitating</i>	<i>Useful</i>	<i>More limited suitability</i>	<i>Less effective preparation¹</i>	<i>Non-counting</i>
Arabic	Ancient history	Art and design ²	Accounting	Critical thinking ^d
Bengali	Archaeology	Business studies	Anthropology ^{dn}	General studies ^d
Biblical Hebrew ⁿ	Classical civilisation	DT: product design (3-D design)	Applied art and design (double award) ^{*d}	
Biology	Classics	DT: product design (textiles) ⁿ	Applied art and design ^{*d}	
Chemistry	Computing	DT: systems and control technology	Applied business (double award) ^{*d}	
Chinese	Cymraig iaith gyntaf, Welsh first	Drama and theatre studies	Applied business ^{*d}	
Classical Greek	Economics	Electronics	Applied ICT (double award) ^{*d}	
Cymraig ail iaith, Welsh second	Economics and business ^d	Film studies	Applied ICT ^{*d}	
Dutch	English language and literature	ICT ⁴	Applied science (double award) ^{*d}	
English literature	English language	Law	Applied science ^{*d}	
French	Environmental science	Media studies	Citizenship studies ^d	
Further mathematics	Geology	Music technology	Communication and culture ^d	
Geography	Government and politics	Physical education	Creative writing ^{dn}	
German	History of art	World development ^d	Dance	
Greek (modern)	Music		DT: food technology	
Gujarati	Philosophy		Engineering ^{*d}	
History	Psychology		Health and social care (double) ^{*d}	
Human biology ^{5dn}	Religious studies		Health and social care ^{*d}	
Italian	Sociology		Humanities ^{dn}	
Japanese	Statistics		Leisure studies (double award) ^{*d}	
Latin			Leisure studies ^{*d}	
Mathematics			Media: communication and	
Modern Hebrew			Performances studies ^d	

¹ Applied A-levels marked *

² Includes 6 additional endorsements/pathways

³ No entries in England/combined with other subject in National Pupil Database markedⁿ

⁴ Information and communication technology

⁵ To be discontinued marked^d

Panjabi

Performing arts*^d

Persian

Science in society^{dn}

Physics

Travel and tourism
(double award)*^d

Polish

Travel and tourism*^d

Portuguese

Pure mathematics^d

Russian

Spanish

Turkish

Urdu