Hard and soft choices? Subject selection by schools and students

Davies, Peter; Ercolani, Marco

DOI:
10.1080/03054985.2018.1469483
10.1080/03054985.2018.1469483

HARD AND SOFT CHOICES?
SUBJECT SELECTION BY SCHOOLS AND STUDENTS

Peter Davies [a,c]
and
Marco G. Ercolani [b,c,*]

[a] P.Davies1@bham.ac.uk, School of Education, University of Birmingham
[b] M.G.Ercolani@bham.ac.uk, Birmingham Business School, University of Birmingham
[c] Centre for Higher Education Equity and Access, University of Birmingham

Biographies:

Peter Davies is professor of education policy research at the University of Birmingham and affiliated professor at Stockholm University. He is author of 'Paying for education: debating the price of progress' (2018) and 'Markets for schooling: an economic analysis' (with Nick Adnett) (2002).

Dr Marco Ercolani is a senior lecturer in econometrics at the University of Birmingham. He has published research in several academic journals in various areas of social statistics, including: ageing workforces, public transport, sickness absence, stock fund performance, student assessment, student loan repayments, tax evasion and more.

Keywords: Field of study, subject choice, economics of education, grade expectations, economics education, business education

JEL Codes: A21, I21, J24

Abstract

We present an analysis of A-level subject choices at around age 16 for a cohort of students in English schools who completed their studies in 2014. We examined both the National Pupil Database and a unique rich dataset on the subject preferences and subsequent choices between the ages of 16 and 18 (i.e. GCSE and A-level). We found substantive differences between students' preferences and actual choices of 'hard' and 'soft' post-16 subjects (i.e. A-level). These differences were strongly associated with falsification of students' expectations of examination grades taken at age 16 (i.e. GCSE) in the core subjects of English and Mathematics. The sizes of these falsification effects were much larger than other significant associations such as gender, ethnicity and social class. This suggests that subject choices are not rigidly framed by stable individual preferences and they are therefore open to influence from new information, persuasion and opportunities.

Funding details

This work was supported by the Nuffield Foundation under Grant number EDU/39013.

[*] Corresponding Author: Dr Marco G. Ercolani, m.g.ercolani@bham.ac.uk, Department of Economics, Birmingham Business School, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

INTRODUCTION

Subject choices in schools matter for future employment, social mobility and the balance of knowledge and skills available for the economy. This study examines factors associated with choice of 'hard' or 'soft' subjects in the final years of secondary schooling in England. 'Hard' subjects are preferred for admission to study at research intensive Russell Group universities, but it is easier for students to gain high grades in 'soft' subjects. We use a comparison between economics and business studies as a focus for our analysis as one subject is seen as 'hard' and the other is seen as 'soft'. Whilst only the study of Mathematics at school has been shown to be associated with higher future income in the UK and the US (Levine and Zimmerman 1995, Dolton and Vignoles 2002, Arcidiacono 2004, Rose and Betts 2004), the subjects which students choose to study in secondary schools set them on trajectories towards different universities, different degree subjects and different employment prospects (Chevalier 2011). Schools' freedom to choose which subjects to offer and students' freedom to choose which subjects to study varies considerably between countries and within countries over time. This has substantial implications for the role of education in society. For example, in the US, students face choices between bundles of subjects in different curriculum tracks leading to different college and non-college trajectories (Zietz and Joshi 2005).

In the UK, social mobility is associated with subject choice through attendance at 'elite' Russell Group universities. Attending a Russell Group university is associated with achieving employment in high status profession and with higher earnings (Sutton Trust 2004, Boliver 2013, Gregg et al. 2013) but also requires high entrance grades. The Russell Group of universities has published advice to school students about which subjects to study when aged 16-18. These 'preferred' subjects also tend to be subjects which are relatively more difficult (CEM 2008) and this presents students with a strategic dilemma: should they choose easier subjects or subjects preferred by elite universities? Students have to make these decisions whilst uncertain about the grades they are likely to achieve. The same decisions also have important implications for schools in a high stakes examination system in which they are judged by the average grade achieved by students. We add to the existing evidence base through new analysis of a national student database and through a unique data set which includes students' expectations of examination grades at age 16 and their plans for subjects to study before they receive the results of the examinations taken at age 16. These data offer insights into relationships between students' strategies and choices that have not been possible with the data sets used in previous studies or through national data sets on subjects studied in England. We find that, ceteris paribus, males and non-white students (i) are more likely than other students to follow the advice of elite universities and avoid 'soft subjects'. (ii) are highly responsive to falsifications of their expected grades especially in mathematics; and (iii) that the strong bias in favour of males taking economics is strengthened by which schools offer the subject and that this bias is stronger in terms of who actually studies the subject compared with those who expected to take the subject. These and other results suggest a substantial malleability in students' choices which leaves them fairly open to influence through accurate or inaccurate information and the pressures that schools may choose to exert.

The next section reviews the background theory and evidence about the role of schools in shaping students' choice of A-level subjects. This followed by an account of method, results and our conclusions.

I. BACKGROUND

In this section we provide a background to the English education system and the subject choices that students have to make at the start of their final two years of secondary education. We then go on to consider the possible incentives facing schools in assisting students with these subject choices.

Advanced level subject difficulty and individual choice

School exit examinations provide the credentials on which applications to university courses are judged in many countries. In England, these judgements are made on the basis of a tariff system which awards points to grades achieved in different types of examination. Discussion of applications to university is largely conducted in terms of grades achieved in A-level courses for which the 2014 tariff points were: A* (140 points), A (120), B (100), C (80), D (60), E(40). Most students sit examinations in three A-levels. Students maximise their chance of entry into university by maximising their A-level tariff points. As well as enabling greater choice between universities, higher grades enable students to reach selective entrance criteria. For example, courses at Russell Group universities typically required at least grades ABB, i.e. 320 points (Russell Group 2011). Whilst Hussain et al. (2009) reported a small additional wage premium for graduates from elite universities in the UK, Walker and Zhu (2013) found no difference between wage premia to graduates of different types of university.

The problem for each student *i* is to maximise the expected total grade tariff G^e from three *j* A-level grades:

$$\max \sum_{j=1}^{3} G_{ij}^{e}(A_{ij}^{t-1}, SE_{ij}, D_j) \quad (1)$$

where A_{ij}^{t-1} is student *i*'s previous attainment in subject *j*, SE_{ij} is student *i*'s self-efficacy (confidence) in subject *j* and D_j is the difficulty of subject *j* relative to other subjects.

Students will choose subjects in which they have a relative advantage (Davies et al. 2009) with the proviso that all students will be discouraged from choosing subjects in which it is harder to achieve top grades. Coe et al. (2008) reviewed different estimates of the difficulty of school subjects. They found that, at A-level, science and modern foreign languages are relatively hard and that most applied subjects such as communication studies and theatre studies are, relatively, easier. They also reported that these differences have been stable over time. Given that high grades are more difficult to achieve in Economics than Business Studies there has been an incentive for students to prefer Business Studies (Bachan and Barrow 2006).

Furthermore, UK research-intensive universities (Russell Group 2011 pp.22-23) expressed their preference for some pre-university (e.g. A-level) subjects rather than others. They split subjects into three groups; hard facilitating, hard non-facilitating, and soft. Hard subjects were defined as "suitable preparation for university". The Russell Group (2013, pp.24-28) subsequently dropped the terms ‘hard’ and ‘soft’ but retained the same three-fold classification. Furthermore, Johnston et al. (2014) found that Economics has gradually become a de-facto “elite subject for elite UK universities”. We will use the language of the 2011 publication since this version was the one available to schools at the time of our study:

Hard traditional (‘facilitating’): Biology, Chemistry, English Literature, Geography, History, Mathematics, Further Mathematics, Foreign Languages (Classical and Modern), Physics.

Hard non-traditional (‘not facilitating’): Classics, Computer Science, Economics, Law, Music, Non-European Languages, Other Science, Philosophy and Religion, Psychology.

Soft: Art, Beauty, Business Studies, Child Development, Design and Technology, Health and Social Care, Media Studies, Performing Arts, Photography, Physical Education, Sociology, Study Skills, Travel and Tourism.

Consequently, if $\sum_{j=1}^{3} G_{ij}^{e} \geq 320$, a student will maximise their university choice by studying at least two ‘facilitating’ subjects (i.e. hard traditional) if this does not prejudice their expectation of achieving a total tariff of at least 320. Otherwise they should simply aim to maximise their grades, with the implication that on average they are less likely than other students to choose ‘facilitating’ subjects because of their higher relative difficulty. In July 2012 The Department for Education (2014) announced that it would be including a measure of school performance in terms of the proportion of students achieving grades AAB in ‘facilitating’ subjects. This introduced a strong incentive for schools to encourage high achieving students to opt for facilitating subjects. This announcement came after the students in this sample had declared their intentions and after schools had determined their curriculum for 2012/13.

We selected our control variables on the basis of theories and evidence regarding the framing of individual choice in relation to gender, socio-economic background and ethnicity. Eccles’ (1994) Expectancy Value Theory identifies interactions between gender and self-efficacy in the context of occupational expectations. Breen and Goldthorpe’s (1997) Relative Risk Aversion Theory predicts that students from high status socio-economic backgrounds will want to match their parents’ occupational status through studying high status subjects. Evidence on this expected social class effect is mixed (cf. Davies and Guppy 1997, Gill and

Bell 2013). In addition, non-white students in the UK appear more likely to aim for high status subjects (Davies et al. 2013). Therefore, we would expect to find that males, students' from high status socio-economic backgrounds and non-white students in England would be more likely to aim to study high status 'hard' subjects.

School effects in subject choice

According to a review by Jin et al. (2013) relatively little of the literature on subject choice has considered school effects although international comparisons suggest that the school attended frames occupational aspirations in differentiated school systems (Buchmann and Dalton 2002). Schools may influence subject choices through the subjects they offer and the advice they give to students. If all schools offered the same advice (e.g. reflecting gender bias in subject selection), then associations between pupil characteristics and subject choice would be affected by schools in ways that statistical analysis would attribute to pupil characteristics. This section concentrates on observable variation between schools operating through a private versus state school effect and through competition between state schools. The chief mechanisms examined here are the positioning of schools in relation to parental aspirations, and the capacity of schools to offer a range of subjects with a sustainable class size.

A difference between private schools and state schools arises from schools' objectives. Private schools focus on the proportion of their students who progress to elite universities (Dunne et al. 2013, Jones 2013) whilst state schools focus on the GCSE and A-level examination grades achieved by their students (Davies et al 2002, Wilson et al. 2006). A relatively high proportion (two thirds) of privately educated students attend elite universities in England (Sutton Trust 2004, Mangan et al. 2010, Gregg et al. 2013), which according to Dunne et al. (2013) and Boliver (2013), reflects the focus of careers guidance in private schools. We interpret this focus as a reflection of what parents are buying when they send their children to private schools: an increased likelihood that their child will enter a well-paid profession on the basis of the advantages conferred by graduating from an elite university (Green et al. 2012, Hussain et al. 2009, McKnight et al. 2002). In order to secure a place at an elite university in the UK it is important to choose some subjects rather than others in the final years of schooling. State schools, which are more focused on between-school competition driven by league tables, face stronger incentives to encourage moderate to low achieving students to study soft subjects. Bachan and Barrow (2006) interpret the relative popularity of Business Studies over Economics in this light.

Researchers in the UK and elsewhere (e.g. Fullarton and Ainley 2000) have consistently found a strong positive, bivariate association between attending private schools and studying traditional, 'hard' subjects. However, at the very least, we might expect this association to be weakened given the inclusion of controls for socioeconomic status, school grades and attendance at private schools. Gill and Bell (2013) analysed factors associated with choosing to study Physics at A-level in England. They found that attending a private school was positively associated with girls choosing physics and negatively associated with boys choosing physics. Using data from the UK Universities and Colleges Admissions Service (UCAS), Boliver (2013) found that students educated in private schools were twice as likely as state school students to apply to an elite university, after controlling for A-level grades. When she added a control for grades in 'facilitating' subjects this private school effect is moderated (the odds ratio of state/private schools increases from 0.48/1 to 0.58/1).

The second process underpinning school effects is school behaviour in the context of quasi-markets. Although competition between schools is frequently focused on when students transfer from primary to secondary schools at age 11, there is also substantial movement of students between schools at age 16 when they have finished compulsory schooling and choose whether to continue schooling for a further two years (Mangan et al. 2001). Adnett and Davies (2000) developed an account of the predictions of economic theory for schools' curriculum design in local markets. They noted that local schooling markets are usually characterised by oligopoly, where the effect on a school of any changes it makes to its curriculum depend on the responses of competitors. Standard analysis of oligopolistic markets suggests that schools' interests may not be aligned with students' interests and that schools near the top of local hierarchies face weak incentives to innovate. Conversely, whilst schools near the bottom of local hierarchies have strong incentives to innovate, they are held back by sunk human capital (the range of expertise of their current staff), the effect of falling enrolment on their capacity to change and their desire to maintain the breadth of the curriculum which fits with their existing conception of the curriculum a school should offer. Follow up studies (Davies et al. 2002, 2003) supported these predictions, with a gradual emergence of curriculum innovation by schools which would be judged as 'less successful' in terms of local league tables and enrolment. Compared with 'more successful' schools these less successful schools had higher proportions of students from lower socio-economic groups. Innovation in these less successful schools tended to be in the form of applied and vocational courses which were seen as more attractive to students and more useful preparation for the futures which these schools envisaged for their students.

II. METHOD

Our research question was:

> How are gender, home background, ethnicity, prior attainment in school, and type of educational institution attended by 16-18 year-olds in England related to the likelihood of choosing hard or soft subjects?

In addition to looking at hard and soft choices as a whole, we focus on choice of business studies or economics. The Russell Group of universities has categorised the former as a 'soft' subject and the latter as a 'hard' (but 'not facilitating') subject. These two subjects are sufficiently close in content to share the same subject association for teachers (the Economics Business and Enterprise Association) and they were also offered jointly as an 'Economics and Business' advanced level. These similarities make it more likely that difference in choice patterns are attributable to difference in subject difficulty and status. We use data from a National Pupil Database (NPD) which reveals national patterns of subject choice by all students aged 16 to 18 for each of the years 2012/13, 2013/14, 2014/15. We also use a rich set of pupil level data which we collected through surveys for a sample of students completing 16-18 education in 2013/14. Our focus is on students who studied at least one A-level subject.

Analysis of national data on A-level entries

From the NPD we use variables measuring: gender; prior attainment at age 16 and type of institution attended. We do not use the measure of social disadvantage (eligibility for free school meals, FSM) provided in this data set. Nor do we use the indicator of ethnicity. This is because these variables were only included in the national dataset for students at age 18 from 2014/15 and we were seeking a comparison with our sample of students completing in 2013/14. We did gain access to a national data source on FSM for 16-18 students, but this had 80% missing data. Moreover, FSM distinguishes rather crudely between the social backgrounds of students enrolling for A-levels. Therefore, we use the national data to provide a general context in which to interpret our survey data whilst also offering a broad picture of relationships between choices, gender, prior attainment and institutional type.

We examined associations in the national data with the likelihood of each of studying: (i) two or more hard subjects; (ii) two or more hard subjects excluding economics and business studies; (iii) economics; (iv) business studies. The analyses of these binary variables were carried out using logistic regression which can be summarised maximising the likelihood function that best fits this equation:

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \frac{e^{\mathbf{x}_i \beta}}{1 + e^{\mathbf{x}_i \beta}} \qquad (2)$$

where *y* is the dependent variable observed on each individual *i,* and x is a vector of regressors. Our earlier analysis suggested a non-linear relationship between prior attainment and the likelihood of studying hard subjects. Increasing prior attainment makes little difference to the likelihood of choosing a hard subject until the student and the school become confident that the student should achieve a top grade. Then there will be a steep rise in probability which tails off beyond the threshold point. Therefore, we included prior attainment squared and prior attainment cubed alongside prior attainment. In equation (2) β is a vector of coefficients and *e* is the natural exponential. Marginal effects for each variable are reported for ease of interpretation. Each marginal effect measures the probability of the outcome switching from 0 to 1 as a result of a one-unit change in that regressor. Marginal effects on the continuous regressors (prior attainment) measure the slope of the probability function with respect to that regressor. The computations are carried out in Stata 13 using the commands -logit- and -margins- (Greene 2012, StataCorp. 2013).

Summary of the survey data

We combined data from three sources to create a unique sample of English schoolchildren before and after their transition at age 16 from basic (GCSE) compulsory to advanced (A-level) voluntary education. These three sources were: (i) a survey of students' expectations at age 15/16 (which provided a rich range of student characteristics); (ii) school reports of subjects studied by these students when aged 17; and (iii) examination grades when aged 16 from the NPD. We use these data to provide a more detailed picture of patterns of choice within a sub-set of institutional types: state and private secondary schools enrolling students up to age 18 and with at least 100 students in the 16-18 age group ('the sixth form'). By using our evidence of difference between students' expected and actual grades at age 16 and difference between the subjects they intended to study and the subjects they actually studied we are able to probe more deeply than has been possible in previous studies. Below we describe the design of the survey, how data from various sources were combined, and the resulting dataset used in our analysis along with summary statistics.

Sampling for the survey data

We approached schools within a large diverse geographical area[1] on a random basis subject to criteria designed to yield a sample that limited the degree of school level variation and facilitated comparison between strategies towards subject choice at age 16. Our criteria for inclusion in the list of schools to be approached were as follows. First, since we were gathering data from students when aged 15-16 we needed sampling criteria that would minimise the proportion of students who would fail to achieve minimum grades at age 16 for A-level study. Second, we were seeking a sample which would include a large proportion of students who had realistic prospects of achieving grades necessary for entry to an elite university. Third, we excluded schools with small sixth forms to reduce the scope for students' decisions to be shaped by schools only offering a few A-level subjects (schools with fewer than 100 students in the sixth form accounted for only 5% of total sixth form enrolment by state and private schools). Fourth, we excluded schools which only offered schooling up to age 16, since patterns of choice and relationships between plans and outcomes could be quite different when students change institution at age 16 (Mangan et al. 2001). Fifth, given the attention paid to private schools in trajectories towards elite universities, we wanted to include a sufficient proportion of private schools to allow a comparison between strategies towards subject choice in the two sectors. Finally, comparisons of state and private schools have to distinguish between consequences of peer group characteristics and consequences of school type. Whilst we are able to take account of a wide range of pupil characteristics (at individual and school level) we also aimed for a sample of state schools which tended to attract (through relatively absolute high examination grades) parents who were more active in their choice of school. By selecting state and private schools from the same broad geographical area helps minimise the unobserved variation in social characteristics that affects student performance and thus mitigates the omitted variable bias. These positive school 'neighbourhood' or 'spillover' effects have been widely documented in England (e.g. Nicoletti and Rabe 2013) and elsewhere (e.g. Böhlmark and Lindahl 2015, albeit for state-funded private schools).

This yielded a total of 958 state schools and 195 private schools. We created two (state and private) randomized lists of schools and invited schools to participate in the order of each list until 50 schools were recruited. We stratified the sample to include 20 private schools to enable a good comparison between practice in the private and state sectors. Due to attrition our final sample consisted of 48 schools, 19 of which were private (6 of which accepted 'boarders' as well as 'day pupils').

Table 1 shows that the students attending schools in our survey sample had relatively high average attainment. The pattern of choice in these schools (where aiming for a Russell Group university was realistic for a relatively high proportion of students) may well be different from choice patterns in schools with a lower proportion of high attaining students.

Table 1 about here

We started by surveying students, aged 15 to 16, in their final year of compulsory education (GCSE examinations). Over 95% of students in the initial survey sample expected to gain grades at age 16 which are regarded as minimum entry levels for university (at least grades C in GCSE Mathematics and English). Roughly half of the students in our sample (52% for English and 56% for mathematics) expected either a grade A* or a grade A in GCSE. That is, our school selection criteria had the effect of generating a suitable sample for comparing strategies in subject selection at schools serving relatively high achieving students. The students in our state schools had lower grade expectations and were less likely to have graduate, professional, parents than the private schools in our sample. However, the academic achievements and socio-economic backgrounds of our state school students were more similar to the private school students in our sample than they were to the averages for all state school students in England.

To check whether our data might be biased by the unwillingness of some schools to participate in the initial survey we compared the private and state schools which agreed to take part in the study with the schools which chose not participate. In total we contacted 189 schools by the time 50 schools firmly committed themselves to participating in the project. We compared means for: School size, numbers of students in the sixth form, % of students gaining 5 grades A* to C at GCSE, value added performance between ages 16 and 18, % of students eligible for free school meals and the index of multiple deprivation of the schools' students postcodes. We found one difference for the state schools: schools which agreed to participate had slightly lower recorded value added scores (991 compared to 1004). We also compared the 19 private schools in our study with all of the other 174 private schools which met the criteria for possible inclusion in the study. We found no significant differences.

Since we match the survey data with information provided by the schools on actual subjects subsequently studied, our analysis is restricted to students who stayed at the same school to study A-level subjects. As expected, students who did not stay at the school to study advanced level subjects had significantly lower GCSE grades in Mathematics (by just over one grade) and English (by roughly one grade). They also had lower cultural capital (just under half a standard deviation) and were less likely to have a father in a professional occupation. However, and crucially for our study, there was no association between staying in the same school and the intention to (i) study hard or soft subjects or (ii) study either Economics or Business Studies.

The Survey Data

Our data on students' characteristics and expectations are drawn chiefly from our survey. We use several indicators of socioeconomic status. These comprised four dummy variables: mother in a professional occupation, father in a professional occupation, graduate mother and graduate father, and one continuous variable: cultural capital. As indicated by Sullivan et al. (2014), the simultaneous inclusion of all these explanatory background characteristics is important. The inclusion of parental background is clearly important to avoid confounding this with school effects (see Hobbs 2016 for a detailed decomposition and Marks 2015 for the bias on school effects when excluding background variables). Parental status, be it educational background or professional employment is widely acknowledged as a positive force on children's educational outcomes primarily through parental nurturing rather than through any direct material rewards, see Darolia and Wydick (2011). Furthermore, according to Taylor and Rampino (2014), this parental effect is stronger during economic downturns such as the one during our survey period. Cultural capital has been frequently cited (e.g. DiMaggio 1982, Noble and Davies 2009) as an important factor in students' higher education choices. Our cultural capital items were drawn from three previous studies (Evans et al. 2010, Noble and Davies 2009, Tramonte and Wilms 2010) and yielded a maximum score of 61 (see Davies et al. 2014 for details of the measure). We normalised this variable to a mean of zero and standard deviation of one.

We also asked students to predict their GCSE examination grades at age 16 and to indicate their preferences for A-level subjects to study in the following academic year. We use students' GCSE grade expectations as indicators of their beliefs about the likelihood that they would reach the 320 A-level points threshold. Furthermore, we asked students for permission to link their data with information in the National Pupil Database which provided their actual

GCSE results, eligibility for free school meals, gender, ethnicity and later their actual GCSE results. We use two variables to capture school-level differences. On the basis of previous research we expected that attending a private school would be positively associated with intending to study and actually studying ‘hard’ subjects. We also expected that students attending schools with a high proportion of high achieving students would be more likely to study hard subjects. This expectation is based on our reading of curriculum effects of between-school competition. We measure this factor in a peer effect calculated as the average grade points at advanced level achieved by students attending each school. We normalised this variable to a mean of zero and standard deviation of one.

In Tables 2, 3 and 4 we report the sample characteristics for the full-cross-section dataset, including students with missing observations among the explanatory variables. These data are used in the main analysis. In Appendix Tables A1, A2 and A3 we report the sample characteristics for the ‘complete-case cross-section’ dataset where students with missing values have been removed. These data are used in the supplementary analysis.

Table 2 summarises intentions to study, and actual choice of, Business Studies or Economics. The table uses responses to a question asking students to indicate how likely they were to study each of a range of advanced level subjects. A Likert scale response format (‘Definitely not’, ‘Unlikely’, ‘Possible’, ‘Likely’, ‘Very Likely’) was transformed into a dummy variable with 1= ‘Very Likely’ or ‘Likely’,

Table 3 presents summary statistics for our categorical variable that captures ‘hard’ versus ‘soft’ A-levels taken, and the Business Studies versus Economics versus neither split. This categorical variable is designed to capture a sufficient number of respondents in each category to carry out a reasonable empirical analysis. Details of how we define this categorical variable are explained in the Appendix. In Table 3 if we combine categories 3 to 7 we see that about 30% of students take either Business Studies and/or Economics at A-level.

Table 4 presents the summary statistics for the explanatory variables. In line with previous research (Chevalier et al. 2009, Hossain and Tsigaris 2012), students in our sample tended to be over-optimistic in the grade expectations, especially in English. We have focused on

GCSE outcomes in English and Mathematics because good performance in these affects opportunities in later life, such as entry into University (and more detailed analyses by subject-area add only a little to predictive power, Benton, 2015).

The peer effect and the cultural capital variables are normalised to have a mean of zero and a variance of one.

Tables 2-4 about here

Model estimators and imputation method

Initially, we (separately) model four binary dependent variables on whether each student intends to, or actually, studies A-level Business Studies or Economics. We then go on to model a single categorical variable that simultaneously captures the overall difficulty of all the A-levels each student has chosen and the combination of choosing A-levels in Business Studies and Economics. Finally, we model a simplified binary version of the categorical variable which permits us to include school fixed effects in the estimates.

The analysis of the binary dependent variables summarized in Table 2 uses the same method of logistic regression which was used to analyse the NPD data, albeit with a richer set of explanatory variables.

To take account of interdependence between subject choices (Johnes 2005) we then analysed the categorical dependent variables summarized in Table 3 using multinomial logistic regression by jointly maximising the joint likelihood function that best fits this set of equations:

$$\Pr(y_i = 1 | \mathbf{x}_i) = \frac{1}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 2 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_2}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 3 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_3}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 4 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_4}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 5 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_5}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 6 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_6}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$
$$\Pr(y_i = 7 | \mathbf{x}_i) = \frac{e^{\mathbf{x}\beta_7}}{1 + e^{\mathbf{x}\beta_2} + e^{\mathbf{x}\beta_3} + e^{\mathbf{x}\beta_4} + e^{\mathbf{x}\beta_5} + e^{\mathbf{x}\beta_6} + e^{\mathbf{x}\beta_7}}$$

where we have arbitrarily used the first equation, and therefore the first outcome, as the reference category. The subscript on each β denotes the equation number. As in the logit, we use the software to calculate all the marginal effects. In each equation the marginal effect is for that one outcome relative to all the other outcomes. One feature of the marginal effects is that these can also be calculated for the reference category, outcome 1 in our case. As a consistency check, we can see that if we take any one regressor and add up (horizontally in the tables) the marginal effects across all seven equations the total should equal zero because all the marginal effects cancel out. The computations were carried out in Stata 13 using the commands -mlogit- and -margins-.

The binary and multinomial analyses described above are carried out using *multiple imputation* (Rubin, 1987, 1996) which allows us to recover missing observations among the explanatory variables. The main reason for doing this is to ensure the results are robust to non-response bias. Another reason for the multiple imputation is simply to increase the sample size. Our operationalization of multiple imputation is in three stages. In the first stage we generate the imputed observations by using linear models on the available observations to generate 30 simulated observations for each missing observation. In effect we are generating a ‘distribution’ of observations for each missing one. In the second stage we estimate the models, either the binary logits or multinomial logits, using appropriate procedures that account the fact that additional observations have been simulated and therefore appropriate weights have to be applied in the regression and adjustments need to be made to the degrees of freedom when calculating the various diagnostic tests. In the final third stage marginal effects are calculated, again, taking account that additional observations have been simulated and making the appropriate adjustments. All three stages of the multiple imputation estimation are carried out in Stata 13 using the multiple imputation commands that are prefixed by -mi-. In stage one the missing data are imputed using the command `mi impute mvn $VarsWithGaps = $VarsNoGaps, add(30)` where the $ prefixes pre-defined variable lists. In the second stage, for example, the logit model is estimated using the command `mi estimate: logit y $RHS` where `y` is the binary dependent variable and `$RHS` is the pre-defined list of explanatory variables. In the final third stage marginal effects are calculated using the command `mimrgns, dydx(*) predict(pr).` `mimrgns` is a user-written alternative by Klein (2014) to Stata's `margins` command that works following multiple imputation `mi`.

III. RESULTS

Analysis of the national dataset

We begin with analyses of the national data for examination entries at age 18. Table 5 reports descriptive data for three cohorts.

Table 5 about here

The total samples for 2012/13 and 2013/14 in Table 5 refer to any student who was examined for any 'Level 3' (A level or equivalent) subject. In 2014/15 the dataset also included students who only studied courses at level 2 (equivalent to GCSE) or below. There was a small increase over these three years in the proportion of students studying 2 or more hard, 'facilitating' subjects ($p<.001$, Chi-squared test). There was also a small increase in the proportion of females studying 1 or more A levels ($p<.001$, Chi-squared test). The likelihood of studying at least one A-level varied hugely according to the type of institution attended.

The coefficients on prior attainment variables confirm the expectation of a non-linear, cubic, relation between prior attainment and the pattern of choice. This is indicated by the coefficients on the 'Prior attainment age 16 (total points)/100' and its square and cube. Table 3 also shows that males and students attending private schools were substantially more likely than others to be studying 2 or more, hard, facilitating subjects. Table 6 also shows that students in the schools included in our survey sample were more likely than other students to be studying 2 or more, hard, facilitating subjects. We repeated the regressions for 2012/13 and for 2014/15 and found very similar patterns.

Table 6 about here

Binary logit regressions on our survey data

We now turn to our unique data set. Tables 7 and 8 present logit model estimates for the full survey dataset with missing observations recovered by multiple imputation. The four binary dependent variables summarised in Table 2. Diagnostics in tables 7 and 8 suggest that the multiple imputation estimates are effective at recovering observations[2] and comparison to estimates in Appendix tables A4 and A5 (without imputation) confirm this.[3] Columns (1) to (3) of each table report Logit regressions for all schools whilst columns (4) to (6) restrict the analysis to schools offering Business Studies (Table 7) or Economics (Table 8). In our complete sample of schools the number of students declaring they were likely or definitely intending to study Economics was higher than the number declaring they were likely or definitely intending to study Business Studies. However, Business Studies was offered more often than Economics as an option. Actual performance in GCSE Mathematics and English is omitted in Tables 7 and 8 to avoid multicollinearity with the variables that measure the difference between actual and expected performance.

Tables 7 and 8 about here

The associations revealed in tables 7 and 8 between choosing Business Studies or Economics and gender, ethnicity and socioeconomic background are similar to those found in previous studies (e.g. Dynan and Rouse 1997, Bachan and Barrow 2006, Davies et al. 2009). Choosing

Economics or Business Studies was negatively associated with being white whilst ethnicity associations with socio-economic background were generally weak. Economics was much more likely to be studied by males but there was a weak gender difference for Business Studies.

We now concentrate on the insights offered from the comparisons unique to this study. Achieving better GCSE grades than expected in Mathematics or English was associated with reduced likelihood of choosing Business Studies (-2.3% to -3.6%). Achieving a better than expected grade in GCSE Mathematics was positively associated with studying Economics (2.6% to 3.8%). This is consistent with our expectation that students expecting high A-level grades will be less likely to study a soft subject. However, it could suggest that students' beliefs about their likely A-level grades change substantially in the light of grades achieved at GCSE. Alternatively, the negative association between unexpected achievement and Business Studies could reflect the way in which options are framed by schools. Students achieving a higher GCSE grade in Mathematics could pass a threshold set by the school for certain subjects or the school might encourage the student to consider a more difficult subject.

We found a small negative association at student level between attending a state school and the intention to study Business Studies (-2.4%) or the actual choice of Business Studies (-6.2%). Although classified as a soft subject, Business Studies was first developed as a school subject in private schools (Davies 1999) and continues to be offered as an option aimed at lower achieving students. Students attending state schools were much less likely to intend to (-7.0%) or actually (-7.6%) study Economics.

Students attending schools with higher achieving peers ('Peer Effect') were less likely to study Business Studies and more likely to study Economics. The negative associations for Business Studies were stronger in actual choices than in intentions (Table 7). Students attending schools that offered Economics were more likely to intend to study the subject if they had more able peers (Table 8). However, this association disappeared in their actual choices. These changes could reflect adjustment in students' choices as they became more aware of the intentions of their peers. Alternatively, it could reflect encouragement in high achieving schools to choose other subjects and encouragement in lower achieving schools to choose Business Studies.

Multinomial logit regressions on the survey data

In this section we report the joint decisions to study Business Studies or Economics in the context of other ‘hard’ or ‘soft’ A-level subjects. This allows us to examine whether the differences observed in choices of Economics or Business Studies were reflected in patterns of choice between hard and soft subjects and also allows us to examine how choice of Economics or Business Studies was framed by other choices of hard and soft subjects. The results are presented in Table 9 which reports the marginal effects from imputed multinomial Logit regressions for the dependent variable ‘A-Level Subject Combination’ defined in Table 3. The Appendix results in Table A6, and the diagnostics in Table 9, confirm that the multiple imputations seem effective in recovering the missing observations.

Table 9 about here

The results in Table 9 (and A6) show that students’ expectations of GCSE grades in Mathematics were strongly associated (in the expected direction) with choices of hard and soft subjects[4]. In addition, the associations between unanticipated high or low grades in Mathematics or English and subject choice was much stronger for combinations of hard or soft subjects than for either Business Studies or Economics.. Students who achieved higher than expected grades in Mathematics tended to switch to studying a smaller percentage of soft subjects[5]. This result was confirmed by analysing a newly constructed student-level variable Gap_i:

$$\text{Gap}_i = [\text{\% of soft subjects studied}]_i - [\text{\% of soft subjects likely to or definitely study}]_i$$

Using an OLS regression with Gap_i as a dependent variable (not reported) we found that higher than expected grades in Mathematics were associated with a smaller than expected percentage of soft subjects studied ($p<.001$, $\beta = -2.7$). This coefficient was only marginally reduced when a range of pupil and school characteristics were added to the regression.

Females and white students were more likely to study combinations of soft subjects. Students with graduate mothers or graduate fathers were more likely to study combinations of hard subjects but not Business Studies or Economics. These results are in line with predictions of the theories of Expectancy Value and Relative Risk Aversion and reinforce concerns about the potential for inequalities to be perpetuated through educational choices.

Turning to school level relationships, the ratio of 'hard' subjects to 'soft' subjects offered by schools was in line with our expectations. Private schools and schools with higher average A-level point scores offered fewer 'soft' subjects relative to 'hard' subjects. These two variables accounted for 50% of the variance in the ratio of 'hard' to 'soft' subjects on offer. However, the results on school-level variables in Table 9 show that, controlling for other factors, students attending state schools were more likely to study combinations of hard subjects and less likely than students in private schools to study combinations of soft subjects. As a robustness check to this curious result we conducted a further analysis (not reported) which excluded the variables related to student achievement at individual and school level. Excluding these variables made no significant difference to the results other than increasing the significance of the 'Peer Effect' variable but this is likely a simple consequence of omitted variable bias resulting from the correlation between individual and school achievement. Nonetheless, our expectation had been that students at private schools would be more likely to study combinations of hard subjects. Given our sampling criteria, state schools with high proportions of high achieving students are over-represented in our sample and this may explain why our results show less difference between state and private schools than might be expected from a nationally representative sample. In contrast to the binomial results for Business Studies and Economics, we found no significant peer effect for combinations of hard and soft subjects which excluded Business Studies and Economics.

We now turn to the framing of choices of Business Studies and Economics within combinations of hard or soft subjects. The positive associations between males and choosing Economics and between being non-white and choosing Economics were stronger with combinations of hard subjects than with combinations of soft subjects. In a similar vein, the positive associations between choosing Economics and expected Mathematics GCSE grade and the negative association between choosing Economics and expected GCSE grade in English were only significant in the context of a combination of hard subjects. Business Studies was more likely to be chosen by non-white students with combinations of hard subjects whilst a GCSE grade above expectations was associated with a lower likelihood of choosing Business Studies in combination with hard subjects. Surprisingly, students in state schools were less likely to choose Business Studies in combination with soft subjects, although this association disappears when achievement variables are omitted (Table 9).

Linear Probability Models on the survey data

In this final sub-section we briefly discuss four Linear Probability Model (LPM) estimates to verify the robustness of the results to the inclusion of school-level fixed effects. To achieve this[6] we generated a binary dependent variable '2+ hard traditional A-levels' summarized at the bottom of Table 3, based on the multinomial dependent variable presented at the top of Table 3. The four resulting regressions are reported in Table 10 and appendix Table 7A.

Table 10 about here

Comparing model (1) to model (2) in Table 10 we see that most demographic variables continue to be significant when we add school fixed effects to the model. The mother's characteristics (Graduate and Professional) become insignificant which suggests that these are highly correlated with the school fixed effects. The other demographic characteristics remain significant but less so, here too showing a degree of correlation with the school fixed effects.

Comparing models (3) and (4) to models (1) and (2) in Table 10 we can see the effect of including expectations of GCSE grades on the importance of the demographic characteristics. Only being Male (positive effect) or White (negative effect) continue to have a significant impact on A-level choices. Family Cultural Capital continues to show a positive effect in model (3) where we allow for GCSE grade expectations but becomes insignificant in model (4) where school fixed effects are also include in the model.

What stands out from Table 10 is that the expectations variables are much more significant and robust in explaining the decision to take ‘hard, traditional’ A-level subjects than the demographic variables when all these are included simultaneously. With the possible exception of being male or white which also remain significant in model (4).

IV. CONCLUSIONS

This study examined the subject choices of English secondary school students in the context of government policy and advice from research intensive universities which has encouraged students to study ‘hard/traditional’ subjects. We found a strong association between high GCSE grades in Mathematics and English and the likelihood of studying ‘hard’ subjects. However, whilst there is a positive association between GCSE Mathematics grade and studying Economics at A-level, the association between studying Economics and GCSE English grade is negative. Our data provide some support for two well-known theories (Expectancy Value Theory and Relative Risk Aversion Theory) which predict that males and students from higher socio-economic backgrounds will be more likely to study ‘hard’ subjects.

Our unique data set also enabled us to examine stability in the relationship between subject preferences expressed when aged 15/16 and actual choices when aged 16/17. We found evidence of substantial switching between hard and soft A-level subjects in response to differences between expected and actual GCSE grades in English and Mathematics. These associations were much stronger for Mathematics than English. This indicates that either students or schools (or both) are willing to put aside their previous predictions of achievement in response to examination grades which are known to include measurement error. These shifts in subject choice can have substantial long-term implications for individual students given the expressed preference of Russell Group universities for ‘hard’ subjects.

School policies on required grade thresholds for studying ‘hard’ A-level subjects might explain the strength of the relationship between unexpected GCSE grades in Mathematics or

English and choice of subject combinations. However, further research would be needed to establish the relative strength of individual and school-level effects. We are, though, able to report several school-level associations. Whilst most schools offer most hard-traditional subjects they supplement these through a range of subjects which have been classified by the Russell Group universities as either 'hard'/'non-facilitating' or 'soft'. Economics is an example of the former and Business Studies is an example of the latter. In our sample, schools that offered economics were more likely to be in the private sector and more likely to have high achieving students. Schools that offered Business Studies were more likely to be in the state sector and more likely to have lower achieving students. However, within this context we found that students in state schools were more likely to be studying 'hard' subjects than students at private schools. This association disappears once expected and actual individual achievement and peer effects are removed. Nonetheless, we had expected to find that students at private schools were more likely than students at state schools to study hard subjects. One possible factor here is the nature of our sample. Our criteria for school selection meant that the state schools in the sample have a high proportion of high achieving students. Another factor is that students attending the private schools in our sample were much more likely to be studying 'hard'/non-facilitating subjects than students at state schools. Turning specifically to Business Studies and Economics we found that students were more likely to study Economics if they were attending a school where the average A-level student achieved high grades whilst students were more likely to study Business Studies if they attended a school where their peers achieved lower A-level grades. These findings extend the rather limited evidence base on school effects in subject choice (Jin et al. 2013).

One way in which government policy could respond to the challenges in a subject choice system is by requiring all students to study a particular range of subjects through, for example, a baccalaureate system. There are two disadvantages of this policy. First, it means that grades will fall (unless there is grade inflation) since some students will no longer be able to select a narrow range of subjects they are best at (Davies et al. 2009). Second, it means that depth of learning will be sacrificed for breadth. Studies of breadth of curriculum choice have, thus far, found no advantage to mixing a broad range of subjects. Indeed, the higher education system is predicated on in-depth study of a few subjects, a model which the Bologna process has encouraged the rest of Europe towards. An alternative would be to change the incentives for schools by making them focus on what students do after leaving school rather than on their examination grades at school. Given that so much of recent education policy has been based on trying to emulate the private sector, it seems odd that this has not been done already.

APPENDIX

The categorical variable: 'A-level subject combination'

We needed criteria by which to define the categorical dependent variable in our analysis in order to capture the proportion of 'hard' or 'soft' subjects taken at A-level and the decision to study Economics and/or Business Studies. Most subjects offered by a school are weak alternatives for each other (with small negative bivariate correlations). However, some subjects (notably in science) are strong complements (with positive bivariate correlations of between 0.3 and 0.4). We model students' choice of subjects using a multinomial function in which we distinguish between choices according to the proportion of 'hard subjects' in students' choice of courses. We follow the three-fold classification of subjects used by the Russell Group of universities (2011, 2013).

Students' choices to study Economics or Business Studies at A-level are framed by this broad categorisation. Table A0 presents an analysis of the other subjects studied by students who had chosen to study either Economics or Business Studies (or both). Compared with students who studied neither Economics nor Business Studies, students who chose Economics also tended to study other hard subjects and students who chose Business Studies also tended to study other soft subjects. The comparisons in Table A0 provide the rationale for placing students in one of the following seven categories according to their subject mix in chosen A-level subjects summarized in Table 2. The categorical variable 'Subject Combination' summarised in Table 2 is the dependent variable in our analysis:

Table A0 - How the choice of Business Studies and Economics is framed by the hard/soft subject distinction

	Students who have studied			
Proportions:	Economics but not Business Studies	Business Studies but not Economics	Economics and Business Studies	Neither Economics nor Business Studies
% of subjects 'hard traditional' (facilitating)	76.3% (26.0)	40.6% (33.5)	25.9% (23.1)	63.7% (29.4)
% of other subjects 'hard but non-traditional'	11.1% (18.0)	24.3% (27.3)	54.3% (23.4)	16.5% (18.5)
% of other subjects 'soft'	12.6% (19.9)	35.2% (31.5)	19.7% (21.5)	19.8% (23.0)
	100%	100%	100%	100%
Observations (2,929 tot.):	503	348	46	2,032

Figures in (parentheses) show standard deviations.

ACKNOWLEDGEMENTS

We acknowledge and thank the Nuffield Foundation for funding the data collection for this project and the University of Birmingham for funding the research and dissemination. We thank seminar participants at the Developments in Economics Education Conference 2015 at the University of Birmingham and at the Work, Pensions and Labour Economics Study Group conference 2015 at the University of Sheffield for their helpful comments. We also thank an anonymous referee for helpful comments and suggestions. All remaining errors are our own.

Ethics approval was obtained via the University of Birmingham under reference number ERN_10-1340 (July 2011). Any opinions expressed herein are those of the authors.

REFERENCES

Adnett, N. & Davies, P. (2000). Competition and Curriculum Diversity in local schooling markets. *Journal of Education Policy* 15(2), 157-167.

Arcidiacono, P. (2004). Ability sorting and the returns to college major, *Journal of Econometrics*, 121, 343-375.

Bachan, R. & Barrow, M. (2006). Modelling curriculum choice at a-level: why is business studies more popular than economics? *International Review of Economics Education*, *5*(2), 58-80.

Benton, T. (2015). Can we do better than using 'mean GCSE grade' to predict future outcomes? An evaluation of General Boosting Models. *Oxford Review of Education*, 41(5), 587-607.

Bohlmark, A. & Lindahl, M. (2015), Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform. *Economica*, 82, 508–551.

Boliver, V. (2013). How fair is access to more prestigious UK universities? *The British Journal of Sociology*, *64*(2), 344-364.

Breen, R. & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Toward a Formal Rational Action Theory. *Rationality and Society*, 9(3), 275–305.

Buchmann & Dalton, B. (2002). Interpersonal influences and educational aspirations in 12 countries: The importance of institutional context. *Sociology of Education,* 75(2), 99-122.

Chevalier A. (2011). Subject choice and earnings of UK graduates, *Economics of Education Review,* 30(6), 1187-1201.

Chevalier, A., Gibbons, S., Thorpe, A., Snell, M. & Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review, 28*(6), 716-727.

Curriculum, Evaluation and Management Centre (2008) *Relative Difficulty of Examinations in Different Subjects*. Durham, CEM Centre. Available online at www.score-education.org/media/3194/relativedifficulty.pdf Accessed 14.06.13

Darolia, R. & Wydick, B. (2011). The Economics of Parenting, Self-esteem and Academic Performance: Theory and a Test. *Economica*, 78, 215–239.

Davies, P.. (1999). The Changing Definition of Business Studies in the School Curriculum. *The Curriculum Journal,* 10(1), 11-28.

Davies, P., Adnett, N. & Mangan, J. (2002). The diversity and dynamics of competition: evidence from two local schooling markets. *Oxford Review of Education*, 28(1),.91-107.

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

Davies, P., Davies, N., Hutton, D., Adnett, N. & Coe, R. (2009). Choosing *in* schools: locating the benefits of specialisation, *Oxford Review of Education,* 35(2),147-167.

Davies, P., Mangan, J., Hughes, A. & Slack, K. (2013) Labour market motivation and students' choice of degree subject. *British Educational Research Journal.* 39(2),361-382.

Davies, P., Qiu, T. & Davies, N. (2014). Cultural and Human Capital, Information and Higher Education Choices, *Journal of Education Policy* 29(6), 804-825.

Davies, S. & Guppy, N. (1997).Fields of study, college selectivity, and student inequalities in higher education. *Social Forces*, 75(4), 1417–1438.

Department for Education (2014). *School and College Performance Tables: Statement of Intent*. London, Department for Education.

DiMaggio, P. (1982). Cultural Capital and School Success: the Impact of Status Culture Participation on the Grades of US High School Students, *American Sociological Review*, 47(2),189-201.

Dolton, P. & Vignoles, A. (2002). The Return on Post–Compulsory School Mathematics Study, *Economica*, 69(273), 113–142.

Dunne, M., King, R. & Ahrens, J. (2014). Applying to higher education: comparisons of independent and state schools. *Studies in Higher Education*, (39)9, 1649-1667.

Dynan, K. E. & Rouse, C. E. (1997).The under-representation of women in economics: a study of undergraduate economics students, *Journal of Economic Education*, 28, 350-368.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly*, *18*(4), 585-609.

Evans, M. D. R., Kelley, J., Sikora, J. & Treiman, D. J. (2010) Family Scholarly Culture and educational success: Books and schooling in 27 nations, *Research in Social Stratification and Mobility*, 28(2), 171-197.

Fullarton, S. & Ainley, J.. (2000). *Subject choice by students in Year 12 in Australian secondary schools. LSAY Research Reports. Longitudinal surveys of Australian youth research report 15*. Available online at http://research.acer.edu.au/lsay_research/13. Accessed 11th October, 2017.

Gill, T. & Bell, J. F. (2013). What Factors Determine the Uptake of A-level Physics? *International Journal of Science Education*, *35*(5), 753-772.

Green, F., Machin, S., Murphy, R. & Zhu, Y. (2012). The Changing Economic Advantage from Private Schools, *Economica*, 79(316), 658-679.

Greene, W. H. (2012). *Econometric Analysis. 7th ed.* Upper Saddle River, NJ: Prentice Hall.

Gregg, P., Milburn, A., Shephard, G., Cleal, P., Attwood, T., Johnston, D., …Williams, C. (2013). *Higher Education: the Fair Access Challenge*. London, Social Mobility and Child Poverty (SMCP) Commission.

Hossai, B. & Tsigaris, P. (2015). Are grade expectations rational? A classroom experiment. *Education Economics* 23(2), 199-212.

Hussain, I., McNally, S. & Telhaj, S. (2009). *University Quality and Graduate Wages in the UK.* LondonSchool of Economics and Political Science, Centre for Economics of Education, Discussion Paper No.0099. London, LSE.

Jin, W., Muriel, A. & Sibieta, L. (2013). *Subject and course choices at age 14 and 16 amongst young people in England: insights from behavioural economics.* Research Report DFE RR-160. London, Department for Education.

Johnes, G. (2005). 'Don't Know Much About History…': Revisiting the Impact of Curriculum on Subsequent Labour Market Outcomes. *Bulletin of Economic Research*, *57*(3), 249-271.

Jones, S. (2013). “Ensure That You Stand Out from the Crowd”: A Corpus-Based Analysis of Personal Statements according to Applicants’ School Type. *Comparative Education Review*, *57*(3), 397-423.

Klein, D. (2014). MIMRGNS: Stata module to run margins after mi estimate, Statistical Software Components S457795, revised 29 Aug 2015. Boston, US, Boston College Department of Economics.

Levine, P. B. & Zimmerman, D. J.. (1995). The Benefit of Additional High School Math and Science Courses for Young Men and Women, *Journal of Business and Economic Statistics*, 13(2), 137-149.

Mangan, J., Davies, P. & Adnett, N. (2001). Movers and stayers: the determinants of post-16 educational choice. *Research in Post-Compulsory Education*, 6(1), 31-50.

Mangan, J., Hughes, A., Davies, P. & Slack, K. (2010). Fair Access: explaining the association between social class and students’ choice of university. *Studies in Higher Education*. 35(3), 335-350.

McKnight, A., Naylor, R. & Smith, J. (2002). *Sheer Class? The extent and sources of variation in the UK Graduate earnings Premium*. LSE STICERD Research Paper No. CASE 054. London, London School of Economics.

Nicoletti, C. & Rabe, B. (2013), Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter? *Economica*, 80(318), 197–218.

Noble, J. & Davies, P. (2009). Cultural capital as an explanation of variation in participation in higher education. *British Journal of Sociology of Education*, 30(5), 591-605.

Perry, T., Davies, P. & Qiu, T. (forthcoming). Great Grade Expectations? The Role of Pupil Expectations in Target Setting, *International Journal of Educational Research.*

Reilly, B. & Bachan, R. (2005). A comparison of A-level performance in economics and business studies: How much more difficult is economics? *Education Economics*, *13*(1), 85-108.

Rose, H. & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, 86(2), 497-513.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, New York, Wiley.

Rubin, D. B. (1996). Multiple imputation after 18#years (with discussion), *Journal of the American Statistical Association*, 91(434), 473—518.

Russell Group (2011). *Informed Choices.* London, The Russell Group of Universities. Available online at www.russellgroup.ac.uk/informed-choices.aspx. Accessed 10.2.12.

Russell Group (2013). *Informed Choices.* London, The Russell Group of Universities. Available online at www.russellgroup.ac.uk/informed-choices.aspx. Accessed 14.6.13.

Smyth, E. & Hanan, C. (2007). School processes and the transition to higher education. *Oxford Review of Education*, 33(2). 175-94.

STATACORP. (2013). *Stata: Release 13. Statistical Software*. College Station, TX: StataCorp LP.

Taylor, M. & Rampino, T. (2014), Educational Aspirations and Attitudes over the Business Cycle. *Economica*, 81(324), 649–673.

The Sutton Trust (2004). *The Missing 3,000: State school students under-represented at leading universities.* London, The Sutton Trust.

Tramonte, L. & Wilms, J. D. (2010). Cultural capital and its effect on educational outcomes, *Economics of Education Review*, 29(.2), 200-213.

Wilson, D., Croxson, B. & Atkinson, A. . (2006). “What gets measured gets done”. Headteachers’ responses to the English Secondary School Performance Management System 1. According to Belh (2003, p. 599), “what gets measured done” is perhaps the most famous aphorism of performance measurement. *Policy Studies*, *27*(2), 153-171.

Zietz, J. & Joshi, P. (2005). Academic choice behavior of high school students: economic rationale and empirical evidence. *Economics of Education Review*, 24(3), 297–308.

Table 1 comparison of survey schools and national averages for attainment and FSM

	State Schools			Private schools		
	In sample	Not in sample	*p*	In sample	Not in sample	*p*
% of pupils eligible for free school meals	17	22	.12	Not available		
% of pupils achieving 5 GCSE grades at A*-C including mathematics and English	71	64	.03	94	85	<.001

Source: Department for Education, School Performance Data, 2011-2012.
p represents the statistical significance in a null hypothesis test that the two means are equal.

Table 2 Summary statistics for the four binary dependent variables in survey data

Variable:	Mean	St. dev.
Intends to study Business Studies A-level	0.159	0.366
Actually studied Business Studies A-level	0.134	0.341
Intends to study Economics A-level	0.206	0.405
Actually studied Economics A-level	0.188	0.390
(2929 Observations)		

Full cross-section dataset, with missing cases, used in estimated models with imputed values.

Table 3 Summary statistics for categorical and simplified binary dependent variables in the raw survey data†

#	A-level subject combination:	Freq.	%	Description on A-levels taken:
1	2+ hard A-levels NoEc NoBu	1,581	54.0	Neither Economics nor Business Studies AND two or more of subjects are hard (i.e. traditional)
2	<2 hard A-levels NoEc NoBu	451	15.4	Neither Economics nor Business Studies AND less than two of subjects are hard
3	2+ hard A-levels & Eco, NoBu	415	14.2	Economics (not Business Studies) AND two or more of subjects are hard
4	<2 hard A-levels & Eco, NoBu	88	3.0	Economics (not Business Studies) AND less than two of other subjects are hard
5	2+ hard A-levels & Bus, NoEc	120	4.1	Business Studies (not Economics) AND 50% or more of subjects are hard
6	<2 hard A-levels & Bus, NoEc	228	7.8	Business Studies (not Economics) AND less than 50% of other subjects are soft
7	Eco & Bus	46	1.6	Economics and Business Studies (sample too small for a hard/soft split)
	Total	2,929	100%	

	Simplified binary variable:	Freq.	%	Description on A-levels taken:
1	2+ hard A-levels taken	2,123	72.5	Studied for two or more hard, traditional A-levels
0	<2 hard A-levels taken	806	27.5	Studied for one or no hard, traditional A-levels
	Total	2,929	100%	

† Full unbalanced cross-section dataset with missing cases used in both the imputed data regressions and the balanced dataset regressions.

Table 4 Summary statistics on control variables in survey data

Variable	Obs.	Unique	Mean	Min.	Max.
Individual's Grades					
Expected GCSE Grade Maths†	2866	11	6.900	3	8
Expected GCSE Grade English†	2855	11	6.756	2	8
Actual - Expected GCSE Grade Maths†	2639	12	-0.083	-4	2
Actual - Expected GCSE Grade English†	2620	13	-0.239	-4	4
School level variables					
State School	2931	2	0.615	0	1
Peer Effect (school average A-level point score normalized to: *N*(0,1))	2931	45	0.000	-2.22	1.44
Demographics					
Male	2912	2	0.510	0	1
White	2907	2	0.745	0	1
Mother Univ. Graduate	2583	2	0.518	0	1
Father Univ. Graduate	2561	2	0.568	0	1
Mother professional	2700	2	0.460	0	1
Father professional	2721	2	0.642	0	1
Family cultural capital, incl. books (normalized to: *N*(1,0))	2776	44	0.000	-3.60	3.07

† GCSE grades are converted to a scale from 8 for an A* down to 2 for an F.
Full cross-section dataset, with missing cases, used in estimated models with imputed values.

Table 5: Summary statistics on all students completing Key Stage 5 in England, 2013-2015 (Source: NPD)

	Whole sample of students			Sub-sample who "Took at least 1 A-level"		
	2012/13	2013/14	2014/15	2012/13	2013/14	2014/15
Mean values for binary dependent variables:						
Took at least 1 A-level	0.374	0.358	0.231	1	1	1
Took 2 or more 'Hard' A-levels	0.146	0.142	0.093	0.389	0.396	0.403
Took 2 or more 'Hard' A-levels but not Business or Economics	0.130	0.126	0.082	0.345	0.352	0.355
Took Business A-level	0.032	0.030	0.019	0.085	0.083	0.083
Took Economics A-level	0.030	0.029	0.021	0.079	0.080	0.090
Mean values for binary explanatory variables:						
State School	0.924	0.926	0.921	0.872	0.871	0.869
Male	0.472	0.470	0.475	0.454	0.451	0.446
Attended 6th form college	0.172	0.171	0.161	0.197	0.194	0.195
Attended Further Ed. College	0.254	0.263	0.249	0.106	0.103	0.093
School participated in our survey:	0.014	0.013	0.013	0.018	0.018	0.018
Observations:	776,918	798,115	1,264,622	290,403	285,369	291,940
Mean values for continuous explanatory variable:						
Prior attainment aged 16 (total points)	520.0	524.4	504.9	545.8	553.4	540.3
Observations:	473,654	493,343	1,264,622	290,403	285,369	291,940

Table 6: Logit regressions on A-level choices (Sample: NPD 2013/14)

	(1) Took 2 or more 'Hard' A-levels	(2) Took 2 or more 'Hard' A-levels but not Business or Economics	(3) Took Business A-level	(4) Took Economics A-level
Male	0.048**	0.027**	0.020**	0.041**
State School	-0.158**	-0.110**	-0.015**	-0.052**
Attended 6th form college	-0.076**	-0.070**	0.004**	-0.004**
Attended F.E. college	-0.343**	-0.320**	-0.066**	-0.093**
School participated in our survey	0.048**	0.035**	0.001	0.004**
Prior attainment age 16 (total points)/100	-0.229**	-0.195**	0.015**	-0.040**
(Prior attainment age 16 (total points)/100)^2	0.068**	0.060**	-0.003**	0.012**
(Prior attainment age 16 (total points)/100)^3	-0.005**	-0.004**	0.000**	-0.001**
Observations	493,343	493,343	493,343	493,343
Pseudo R^2	0.2312	0.2020	0.0500	0.1360

Standard errors on all marginal effects <0.01, not reported

* $p < 0.10$, ** $p < 0.05$

Table 7 Marginal effects from multiply-imputed Logit regressions on intending to or actually studying Business Studies at A-level

	(1)	(2)	(3)	(4)	(5)	(6)
	All schools:			Schools offering Business Studies:		
Took Business Studies?	*Intended to*	*Actually did*	*Actually did*	*Intended to*	*Actually did*	*Actually did*
Expected GCSE Grade Maths	-0.000	-0.037	-0.035	0.001	-0.041	-0.040
	(0.01)	(0.01)**	(0.01)**	(0.01)	(0.01)**	(0.01)**
Expected GCSE Grade English	-0.045	-0.035	-0.046	-0.040	-0.037	-0.049
	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**
Actual - Expected GCSE Grade Maths			-0.023			-0.029
			(0.01)**			(0.01)**
Actual - Expected GCSE Grade English			-0.036			-0.036
			(0.01)**			(0.01)**
State School	-0.024	-0.032	-0.033	-0.041	-0.062	-0.061
	(0.02)	(0.02)*	(0.02)*	(0.02)*	(0.02)**	(0.02)**
Peer Effect	-0.002	-0.030	-0.021	-0.011	-0.024	-0.014
	(0.01)	(0.01)**	(0.01)**	(0.01)	(0.01)**	(0.01)
Male	0.019	0.032	0.023	0.031	0.043	0.033
	(0.01)	(0.01)**	(0.01)*	(0.02)*	(0.02)**	(0.02)**
White	-0.074	-0.045	-0.039	-0.074	-0.058	-0.052
	(0.02)**	(0.01)**	(0.01)**	(0.02)**	(0.02)**	(0.02)**
Mother Univ. Graduate	0.011	0.009	0.013	0.008	0.009	0.014
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Father Univ. Graduate	-0.060	-0.046	-0.043	-0.066	-0.050	-0.047
	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**
Mother professional	-0.027	-0.009	-0.012	-0.030	-0.010	-0.013
	(0.02)*	(0.01)	(0.01)	(0.02)*	(0.02)	(0.02)
Father professional	0.050	0.021	0.025	0.044	0.023	0.026
	(0.02)**	(0.01)	(0.01)*	(0.02)**	(0.02)	(0.02)
Family cultural capital (incl. books)	-0.008	-0.027	-0.025	-0.008	-0.033	-0.030
	(0.01)	(0.01)**	(0.01)**	(0.01)	(0.01)**	(0.01)**
Observations	2929	2929	2929	2399	2399	2399
Imputations	30	30	30	30	30	30
Average relative variance increase	0.0620	0.0694	0.0770	0.0585	0.0688	0.0754
Largest fraction of missing information	0.2074	0.2066	0.2073	0.1833	0.2050	0.2056
F-statistic	7.60**	21.00**	19.45**	6.60**	14.74**	14.04**

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors on marginal effects are reported in (parentheses).
Using full imputed-case dataset, with 30 imputed observations for each missing observation

Table 8 Marginal effects from multiply-imputed Logit regressions on intending to or actually studying Economics at A-level

	(1)	(2)	(3)	(4)	(5)	(6)
	All schools:			Schools offering Economics:		
Took Economics?	*Intended to*	*Actually did*	*Actually did*	*Intended to*	*Actually did*	*Actually did*
Expected GCSE Grade Maths	0.063	0.054	0.062	0.082	0.068	0.080
	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**
Expected GCSE Grade English	-0.032	-0.034	-0.041	-0.038	-0.038	-0.048
	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**
Actual - Expected GCSE Grade Maths			0.026			0.038
			(0.01)**			(0.02)**
Actual - Expected GCSE Grade English			-0.014			-0.018
			(0.01)			(0.01)
State School	-0.076	-0.076	-0.076	-0.072	-0.071	-0.070
	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**
Peer Effect	0.054	0.049	0.048	0.044	0.008	0.005
	(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)	(0.01)
Male	0.102	0.118	0.118	0.130	0.163	0.164
	(0.01)**	(0.01)**	(0.01)**	(0.02)**	(0.02)**	(0.02)**
White	-0.068	-0.034	-0.032	-0.077	-0.050	-0.048
	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**
Mother Univ. Graduate	-0.052	-0.035	-0.035	-0.067	-0.039	-0.040
	(0.02)**	(0.02)*	(0.02)**	(0.02)**	(0.02)*	(0.02)*
Father Univ. Graduate	-0.010	-0.019	-0.020	-0.017	-0.022	-0.024
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Mother professional	0.008	0.014	0.014	0.009	0.011	0.011
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Father professional	0.036	0.008	0.009	0.044	0.003	0.005
	(0.02)**	(0.02)	(0.02)	(0.02)**	(0.02)	(0.02)
Family cultural capital (incl. books)	0.019	-0.003	-0.002	0.020	-0.007	-0.007
	(0.01)**	(0.01)	(0.01)	(0.01)**	(0.01)	(0.01)
Observations	2929	2929	2929	2280	2280	2280
Imputations	30	30	30	30	30	30
Average relative increase in variance	0.0387	0.0319	0.0530	0.0383	0.0328	0.0571
Largest fraction of missing information	0.1233	0.0871	0.1742	0.1168	0.0868	0.1941
F-statistic	28.32**	24.33**	20.31**	19.84**	14.95**	12.71**

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors on marginal effects are reported in (parentheses)
Using full imputed-case dataset, with 30 imputed observations for each missing observation

Table 9:
Marginal effects from a Multinomial Logit regression using unrestricted data†

Dependent variable:	A-level subject combination						
Outcomes:	(1) 2+ hard A-levels NoEc NoBu	(2) <2 hard A-levels NoEc NoBu	(3) 2+ hard A-levels & Eco, NoBu	(4) <2 hard A-levels & Eco, NoBu	(5) 2+ hard A-levels & Bus, NoEc	(6) <2 hard A-levels & Bus, NoEc	(7) Eco & Bus
Expected GCSE Grade Maths	0.049 (0.01)**	-0.073 (0.01)**	0.066 (0.01)**	-0.006 (0.00)	-0.005 (0.00)	-0.035 (0.01)**	0.004 (0.00)
Expected GCSE Grade English	0.102 (0.01)**	-0.023 (0.01)**	-0.027 (0.01)**	-0.006 (0.00)	-0.017 (0.01)**	-0.020 (0.01)**	-0.010 (0.00)**
Actual - Expected GCSE Grade Maths	0.033 (0.02)**	-0.036 (0.01)**	0.034 (0.01)**	-0.003 (0.00)	0.000 (0.01)	-0.026 (0.01)**	-0.002 (0.00)
Actual - Expected GCSE Grade English	0.056 (0.01)**	-0.013 (0.01)	-0.006 (0.01)	-0.004 (0.00)	-0.017 (0.01)**	-0.012 (0.01)*	-0.003 (0.00)
State School	0.090 (0.02)**	0.037 (0.02)*	-0.063 (0.01)**	-0.018 (0.01)**	0.001 (0.01)	-0.058 (0.02)**	0.012 (0.01)
Peer Effect	-0.007 (0.01)	-0.032 (0.01)**	0.047 (0.01)**	0.006 (0.01)	0.003 (0.01)	-0.015 (0.01)*	-0.002 (0.00)
Male	-0.060 (0.02)**	-0.066 (0.01)**	0.088 (0.01)**	0.014 (0.01)**	0.016 (0.01)**	-0.006 (0.01)	0.015 (0.01)**
White	0.018 (0.02)	0.056 (0.02)**	-0.041 (0.01)**	0.011 (0.01)	-0.010 (0.01)	-0.031 (0.01)**	-0.002 (0.01)
Mother Univ. Graduate	0.026 (0.02)	-0.006 (0.02)	-0.027 (0.02)*	-0.005 (0.01)	-0.002 (0.01)	0.016 (0.01)	-0.002 (0.01)
Father Univ. Graduate	0.046 (0.02)**	0.017 (0.02)	-0.011 (0.02)	-0.009 (0.01)	-0.008 (0.01)	-0.031 (0.01)**	-0.004 (0.01)
Mother professional	0.001 (0.02)	-0.004 (0.01)	0.015 (0.01)	-0.002 (0.01)	0.006 (0.01)	-0.017 (0.01)	0.000 (0.01)
Father professional	-0.010 (0.02)	-0.019 (0.01)	0.014 (0.02)	-0.006 (0.01)	0.016 (0.01)*	0.001 (0.01)	0.004 (0.01)
Family cultural capital (incl. books)	0.013 (0.01)	0.008 (0.01)	0.006 (0.01)	-0.002 (0.00)	-0.001 (0.00)	-0.020 (0.01)**	-0.004 (0.00)
Observations	2929						
Imputations	30						
Average relative variance increase	0.0640						
Largest fraction of missing information	0.2644						
F-statistic	11.91**						

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors for marginal effects reported in (parentheses)
† Using full dataset with 30 imputed observations for each missing observation.
Top row abbreviations:
NoEc(onomics A-level), NoBu(siness Studies A-level), Ec(onomics A-level), Bu(siness A-level).

Table 10:
Linear Probability Models (LPM) without and with Fixed Effects (FE), using unrestricted data†

Dependent variable:	2+ hard traditional A-levels			
Model:	(1)	(2)	(3)	(4)
Estimator:	LPM without FE	LPM with FE	LPM without FE	LPM with FE
Expected GCSE Grade Maths			0.161 (0.01)**	0.141 (0.01)**
Expected GCSE Grade English			0.070 (0.01)**	0.067 (0.01)**
Actual - Expected GCSE Grade Maths			0.095 (0.01)**	0.082 (0.01)**
Actual - Expected GCSE Grade English			0.034 (0.01)**	0.041 (0.01)**
State School††				
Peer Effect††				
Male	0.096 (0.02)**	0.045 (0.02)**	0.037 (0.01)**	0.037 (0.02)**
White	-0.104 (0.02)**	-0.035 (0.02)*	-0.036 (0.02)**	-0.031 (0.02)*
Mother Univ. Graduate	0.044 (0.02)**	0.018 (0.02)	-0.007 (0.02)	-0.007 (0.02)
Father Univ. Graduate	0.070 (0.02)**	0.040 (0.02)**	0.025 (0.02)	0.023 (0.02)
Mother professional	0.050 (0.02)**	0.024 (0.02)	0.023 (0.02)	0.018 (0.02)
Father professional	0.094 (0.02)**	0.041 (0.02)**	0.033 (0.02)*	0.023 (0.02)
Family cultural capital (incl. books)	0.084 (0.01)**	0.040 (0.01)**	0.022 (0.01)**	0.013 (0.01)
School Fixed Effects	None	45	None	45
Observations	2929	2929	2929	2929
Imputations	30	30	30	30
Average relative variance increase	0.0679	0.0122	0.0825	0.0253
Largest fraction of missing information	59.33	20.58	118.73	30.17
F-statistic	0.1350	0.1553	0.1585	0.1661

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors for marginal effects reported in (parentheses).
† Using full dataset with 30 imputed observations for each missing observation.
†† Excluded variables due to perfect collinearity in regressions with School fixed effects.

TABLES FOR THE APPENDIX:

Table A1: Summary statistics for the four binary dependent variables, using complete-case cross-section dataset (1983 obs.)

Variable:	Mean	St. dev.
Intends to study Business Studies A-level	0.149	0.357
Actually studied Business Studies A-level	0.115	0.319
Intends to study Economics A-level	0.221	0.415
Actually studied Economics A-level	0.199	0.399

Table A2:
Summary statistics for categorical dependent variable 'A-level subject combination' for complete-case cross-section dataset

#	A-level subject combination:	Freq.	%	Description on A-levels taken:
1	2+ hard A-levels NoEc NoBu	1,059	57.2	Neither Economics nor Business Studies AND two or more of subjects are hard (i.e. traditional)
2	<2 hard A-levels NoEc NoBu	241	13.0	Neither Economics nor Business Studies AND less than two of subjects are hard
3	2+ hard A-levels & Eco, NoBu	290	15.7	Economics (not Business Studies) AND two or more of subjects are hard
4	<2 hard A-levels & Eco, NoBu	50	2.7	Economics (not Business Studies) AND less than two of other subjects are hard
5	2+ hard A-levels & Bus, NoEc	74	4.0	Business Studies (not Economics) AND 50% or more of subjects are hard
6	<2 hard A-levels & Bus, NoEc	110	5.9	Business Studies (not Economics) AND less than 50% of other subjects are soft
7	Eco & Bus	29	1.6	Economics and Business Studies (sample too small for a hard/soft split)
	Total	1,853	100%	

Table A3:
Summary statistics on control variables for complete-case cross-section dataset

Variable	Obs.	Unique	Mean	Min.	Max.
Individual's Grades					
Expected GCSE Grade Maths†	1853	10	7.009	3	8
Expected GCSE Grade English†	1853	10	6.859	3	8
Actual - Expected GCSE Grade Maths†	1853	11	-0.050	-3	2
Actual - Expected GCSE Grade English†	1853	12	-0.206	-3	4
School level variables					
State School	1853	2	0.603	0	1
Peer Effect (school average A-level point score normalized to: *N*(0,1))	1853	44	0.000	-2.26	1.35
Demographics					
Male	1853	2	0.512	0	1
White	1853	2	0.763	0	1
Mother Univ. Graduate	1853	2	0.520	0	1
Father Univ. Graduate	1853	2	0.570	0	1
Mother professional	1853	2	0.484	0	1
Father professional	1853	2	0.671	0	1
Family cultural capital, incl. books (normalized to: *N*(1,0))	1853	43	0.000	-3.47	3.03

† GCSE grades are converted to a scale from 8 for an A* down to 2 for an F.

Table A4
Marginal effects from Logit regressions studying Business Studies A-level, using complete-case cross-section dataset†

	(1)	(2)	(3)	(4)	(5)	(6)
	All schools:			Schools offering Business Studies:		
Took Business Studies?	*Intended to*	*Actually did*	*Actually did*	*Intended to*	*Actually did*	*Actually did*
Expected GCSE Grade Maths	-0.003 (0.01)	-0.033 (0.01)**	-0.033 (0.01)**	0.000 (0.01)	-0.037 (0.01)**	-0.038 (0.01)**
Expected GCSE Grade English	-0.041 (0.01)**	-0.031 (0.01)**	-0.041 (0.01)**	-0.039 (0.01)**	-0.034 (0.01)**	-0.044 (0.01)**
Actual – Expected GCSE Grade Maths			-0.019 (0.01)**			-0.024 (0.01)**
Actual – Expected GCSE Grade English			-0.032 (0.01)**			-0.032 (0.01)**
State School	-0.037 (0.02)*	-0.041 (0.02)*	-0.040 (0.02)*	-0.046 (0.03)*	-0.074 (0.03)**	-0.069 (0.03)**
Peer Effect	-0.014 (0.01)	-0.045 (0.01)**	-0.036 (0.01)**	-0.023 (0.01)	-0.045 (0.01)**	-0.036 (0.01)**
Male	0.011 (0.02)	0.018 (0.01)	0.010 (0.01)	0.018 (0.02)	0.026 (0.02)	0.017 (0.02)
White	-0.066 (0.02)**	-0.050 (0.02)**	-0.047 (0.02)**	-0.065 (0.02)**	-0.062 (0.02)**	-0.060 (0.02)**
Mother Univ. Graduate	0.025 (0.02)	0.019 (0.02)	0.022 (0.02)	0.026 (0.02)	0.020 (0.02)	0.025 (0.02)
Father Univ. Graduate	-0.058 (0.02)**	-0.039 (0.02)**	-0.036 (0.02)**	-0.078 (0.02)**	-0.044 (0.02)**	-0.040 (0.02)*
Mother professional	-0.010 (0.02)	0.008 (0.02)	0.005 (0.02)	-0.008 (0.02)	0.011 (0.02)	0.007 (0.02)
Father professional	0.049 (0.02)**	0.008 (0.02)	0.009 (0.02)	0.047 (0.02)**	0.006 (0.02)	0.009 (0.02)
Family cultural capital (incl. books)	-0.020 (0.01)**	-0.018 (0.01)**	-0.017 (0.01)**	-0.017 (0.01)*	-0.023 (0.01)**	-0.022 (0.01)**
Observations	1853	1853	1853	1478	1478	1478
Pseudo R^2	0.0384	0.1365	0.1512	0.0439	0.0995	0.1119

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors on marginal effects are reported in (parentheses)
† Dataset restricted to all non-missing observations

Table A5:
Marginal effects from Logit regressions studying Economics A-level, using complete-case cross-section dataset†

	(1)	(2)	(3)	(4)	(5)	(6)
	All schools:			Schools offering Economics:		
Took Economics?	*Intended to*	*Actually did*	*Actually did*	*Intended to*	*Actually did*	*Actually did*
Expected GCSE Grade Maths	0.068	0.050	0.060	0.085	0.062	0.077
	(0.01)**	(0.01)**	(0.01)**	(0.02)**	(0.02)**	(0.02)**
Expected GCSE Grade English	-0.033	-0.032	-0.043	-0.040	-0.033	-0.046
	(0.01)**	(0.01)**	(0.01)**	(0.02)**	(0.02)**	(0.02)**
Actual - Expected GCSE Grade Maths			0.030			0.045
			(0.02)*			(0.02)**
Actual - Expected GCSE Grade English			-0.022			-0.025
			(0.01)*			(0.02)
State School	-0.072	-0.065	-0.065	-0.070	-0.052	-0.051
	(0.02)**	(0.02)**	(0.02)**	(0.03)**	(0.03)**	(0.03)**
Peer Effect	0.053	0.042	0.042	0.035	-0.003	-0.005
	(0.01)**	(0.01)**	(0.01)**	(0.02)**	(0.02)	(0.02)
Male	0.106	0.126	0.125	0.131	0.171	0.172
	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.02)**
White	-0.089	-0.045	-0.042	-0.104	-0.059	-0.055
	(0.02)**	(0.02)**	(0.02)**	(0.02)**	(0.03)**	(0.03)**
Mother Univ. Graduate	-0.049	-0.034	-0.035	-0.062	-0.037	-0.040
	(0.02)**	(0.02)	(0.02)	(0.03)**	(0.03)	(0.03)
Father Univ. Graduate	-0.021	-0.013	-0.014	-0.029	-0.013	-0.015
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Mother professional	0.011	0.031	0.029	0.019	0.029	0.028
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Father professional	0.056	0.016	0.016	0.064	0.013	0.015
	(0.02)**	(0.02)	(0.02)	(0.03)**	(0.03)	(0.03)
Family cultural capital (incl. books)	0.013	-0.003	-0.002	0.015	-0.009	-0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	1853	1853	1853	1488	1488	1488
Pseudo R^2	0.1297	0.1011	0.1039	0.1019	0.0691	0.0729

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors on marginal effects are reported in (parentheses)
† Dataset restricted to all non-missing observations

Table A6:
Marginal effects from Multinomial Logit regression, using complete-case cross-section dataset†

Dependent variable:	A-level subject combination						
Outcomes:	(1) Hard NoEc NoBu	(2) Soft NoEc NoBu	(3) Hard & Eco, NoBu	(4) Soft & Eco, NoBu	(5) Hard & Bus, NoEc	(6) Soft & Bus, NoEc	(7) Eco & Bus
Expected GCSE Grade Maths	0.055	-0.070	0.052	-0.005	-0.009	-0.027	0.003
	(0.02)**	(0.01)**	(0.01)**	(0.00)	(0.01)	(0.01)**	(0.00)
Expected GCSE Grade English	0.098	-0.023	-0.026	-0.008	-0.011	-0.019	-0.011
	(0.02)**	(0.01)**	(0.01)**	(0.01)	(0.01)*	(0.01)**	(0.00)**
Expected - Actual GCSE Grade Maths	-0.027	0.033	-0.030	0.003	-0.001	0.020	0.003
	(0.02)	(0.01)**	(0.01)**	(0.01)	(0.01)	(0.01)**	(0.00)
Expected - Actual GCSE Grade English	-0.050	0.002	0.009	0.009	0.015	0.011	0.005
	(0.02)**	(0.01)	(0.01)	(0.01)	(0.01)**	(0.01)	(0.00)
State School	0.105	0.005	-0.055	-0.017	-0.003	-0.055	0.019
	(0.03)**	(0.02)	(0.02)**	(0.01)*	(0.01)	(0.02)**	(0.01)*
Peer Effect	0.012	-0.028	0.046	-0.000	-0.006	-0.023	-0.002
	(0.02)	(0.01)**	(0.01)**	(0.01)	(0.01)	(0.01)**	(0.00)
Male	-0.066	-0.054	0.106	0.007	0.014	-0.017	0.010
	(0.02)**	(0.01)**	(0.02)**	(0.01)	(0.01)	(0.01)	(0.01)*
White	0.042	0.048	-0.038	-0.004	-0.014	-0.027	-0.007
	(0.03)	(0.02)**	(0.02)**	(0.01)	(0.01)	(0.01)**	(0.01)
Mother Univ. Graduate	0.028	-0.010	-0.034	-0.007	-0.008	0.025	0.006
	(0.03)	(0.02)	(0.02)*	(0.01)	(0.01)	(0.01)*	(0.01)
Father Univ. Graduate	0.041	0.001	-0.019	0.007	-0.006	-0.023	-0.001
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)*	(0.01)
Mother professional	-0.022	-0.009	0.033	-0.003	0.021	-0.022	0.002
	(0.02)	(0.02)	(0.02)*	(0.01)	(0.01)**	(0.01)*	(0.01)
Father professional	-0.013	-0.015	0.017	0.000	0.015	-0.002	-0.002
	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Family cultural capital (incl. books)	0.001	0.001	0.001	-0.000	-0.000	-0.002	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)**	(0.00)*
Observations	1853						
Wald χ^2_{78}	667.3**						

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors on marginal effects are reported in (parentheses)
Abbreviations: NoEc(onomics A-level), NoBu(siness Studies A-level), Ec(onomics A-level), Bu(siness A-level)
† Dataset restricted to all non-missing observations

Table A7:
Linear Probability Models (LPM) without and with Fixed Effects (FE), using complete-case cross-section dataset†

Dependent variable:	2+ hard traditional A-levels			
Model:	(1)	(2)	(3)	(4)
Estimator:	LPM without FE	LPM with FE	LPM without FE	LPM with FE
Expected GCSE Grade Maths			0.165 (0.01)**	0.145 (0.01)**
Expected GCSE Grade English			0.075 (0.01)**	0.074 (0.01)**
Actual - Expected GCSE Grade Maths			0.097 (0.01)**	0.084 (0.01)**
Actual - Expected GCSE Grade English			0.029 (0.01)**	0.036 (0.01)**
State School††				
Peer Effect††				
Male	0.092 (0.02)**	0.053 (0.02)**	0.043 (0.02)**	0.049 (0.02)**
White	-0.086 (0.02)**	-0.018 (0.02)	-0.015 (0.02)	-0.014 (0.02)
Mother Univ. Graduate	0.037 (0.02)	0.004 (0.02)	-0.017 (0.02)	-0.021 (0.02)
Father Univ. Graduate	0.065 (0.02)**	0.030 (0.02)	0.014 (0.02)	0.010 (0.02)
Mother professional	0.055 (0.02)**	0.037 (0.02)*	0.032 (0.02)*	0.031 (0.02)*
Father professional	0.080 (0.02)**	0.042 (0.02)**	0.029 (0.02)	0.021 (0.02)
Family cultural capital (incl. books)	0.063 (0.01)**	0.027 (0.01)**	0.012 (0.01)	0.005 (0.01)
School Fixed Effects	None	45	None	45
Observations	1853	1853	1853	1853
R^2	0.1077	0.2569	0.3213	0.3744

Notes: * $p < 10\%$, ** $p < 5\%$. Standard errors for marginal effects reported in (parentheses).
† Dataset restricted to all non-missing observations.
†† Excluded variables due to perfect collinearity in regressions with School fixed effects.

---

[1] This area included all schools teaching pupils aged 16-18 in the post code areas: AL, B, BA. BR, BS, CH, CR, CV, CW, DE, E, EN, GL, HA, HP, IG, KT, L, LE, LU, M, MK, N, NG, NN, NW, OL, OX, RG, RH, RM, SE, SG, SK, SL, SM, ST, SW, TW, UB, W, WA, WD WR, WS, WV. This area was roughly bounded by the cities of Liverpool, Sheffield, London and Bristol.

[2] The 'average relative variance increase' statistics in tables 7 and 8 are all small (though no precise critical values exist) confirming that the missing information did not have a significant impact on the results. The 'largest fraction of missing observations' (LFMI) statistics are also quite small, suggesting that 30 imputations is sufficient. Again, there are no precise critical values for LFMI but the rule of thumb is that for the imputations to be sufficient these should exceed 100 times the LFMI and in all regressions this appears to be true. However, 30 imputations is quite high and earlier attempts, with smaller numbers of imputations, did not pass these

diagnostic tests, particularly when it came to predict the decision to take Business Studies. This need for a large number of imputations might arise from the fact that most of the explanatory variables are missing at least a few observations.

[3] Appendix tables A4 and A5 report the same model estimates using the more traditional 'complete-case cross-section' method with simple case-wise removal of any student with any missing observations. The Appendix results are similar to those in the main body of the text but with about one thousand fewer observations.

[4] We ran a separate multinomial regression omitting the attainment variables (available on request). The associations with student characteristics (particularly socioeconomic status) were strengthened indicating that gender, ethnicity and socioeconomic background influences on subject choice operate partly directly and partly through achievement in school.

[5] We are grateful to a referee who suggested we examine the relationship between average school performance and the difference between a student's actual and expected grade. We found that a small positive association between attending a school with a higher percentage of students gaining 5 or more GCSE grades A*-C and the difference between the student's actual and end expected grade in both mathematics and English. This association was slightly attenuated by controlling for other pupil and school characteristics. Once these controls were added we found that a one standard deviation increase in the proportion of students at a school achieving grades A*-C was positively associated with one sixth of a standard deviation in the difference between actual and expected grade in mathematics and one month of a standard deviation in the difference between actual and expected grade in English. Further analysis of relationships between expected and actual grades in this sample is available in Perry, Davies & Qiu (forthcoming).

[6] It was not possible to include school fixed effects in the multinomial logit regressions. The number of extra parameters (number of schools times number of outcomes) increased the parameter space to such an extent that the maximum likelihood estimator could not achieve convergence.