The socio-economic gap in university drop out

Nattavudh Powdthavee
University of York

Anna Vignoles[1]
Institute of Education, University of London

January 2009

ABSTRACT

In many countries, including the US and the UK, there is ongoing concern about the extent to which young people from poor backgrounds can acquire a university degree. Recent evidence from the UK however, suggests that for a given level of prior achievement in secondary school, a disadvantaged pupil has as much chance of enrolling in a university as a more advantaged student. Yet simply participating in higher education is not sufficient: graduation is as important. This paper therefore investigates whether students from lower socio-economic backgrounds have a higher rate of university drop-out as compared to their wealthier counterparts, once one allows for their differential prior achievement. Using a combination of school and university administrative data sets, we show that there is indeed a sizeable and statistically significant gap in the rate of withdrawal after the first year of university between the most advantaged and disadvantaged English students. This socio-economic gap in university drop-out remains even after allowing for their personal characteristics, prior achievement in secondary school and university characteristics. In the English context at least, this implies that retention in university of disadvantaged students is arguably a more important policy issue than barriers to entry for these students.

Key words: college drop out; college retention; university drop out; prior achievement; socio-economic gap

JEL: I2

[1] Centre for Economics of Education, Institute of Education. Corresponding author: a.vignoles@ioe.ac.uk

1. Introduction

As in the US and in many other countries, the under representation of poorer students in college or university has been an issue of great policy concern in the UK for many decades. Recent evidence for England indicates that the 20% most disadvantaged students are around 6 times less likely to participate in university compared to the 20% most advantaged pupils (HEFCE, 2005). Furthermore, the socio-economic gap in university participation actually widened in the UK in the mid and late 1990s (Blanden and Machin, 2004; Machin and Vignoles, 2004; HEFCE, 2005). As in the US (Cunha et al. 2006), much of the root cause of this inequality is located earlier in the education system. Chowdry, Crawford, Dearden, Goodman and Vignoles (2008) have shown that in England, if a disadvantaged pupil does reach a sufficient level of achievement in secondary school, s/he has a similar chance of going on to university as a more advantaged student.

However, in both the UK and the US, the lower participation rate of poorer students is not the only policy concern. To fully reap the rewards from a university education, poorer students need to complete their degrees. Currently around three quarters of those who start a degree in the UK actually complete it. Although the UK performs well in international league tables of student retention (NAO, 2007), the fact that one quarter of students who enroll in university do not graduate is potentially cause for concern. Specifically, even if disadvantaged students are as likely to participate in higher education for a given level of prior achievement, they may be more likely to drop out than wealthier students. For instance, disadvantaged students may have less complete information about the real costs of studying for a degree, and we know that in the UK poorer students leave university with more debt and are more risk averse in the first place (Pennell and West, 2005). This might mean that students' socio-economic background affects their likelihood of dropping out even if family background does not affect university participation conditional on prior academic achievement (Chowdry et al. 2008, Marcenaro-Gutierrez, Galindo-Rueda and Vignoles, 2007). Certainly in the literature drop out or non-completion has been seen as particularly problematic for students from disadvantaged backgrounds (Dearing,

1997; McGivney, 1996; HEFCE, 1999; Quinn, 2004)[2], and the extent to which drop out does indeed vary by socio-economic background is the focus of this paper.

In this paper we ask whether disadvantaged entrants to college[3] have a higher probability of 'dropping out', given their level of prior achievement. In other words, does disadvantage and poverty mean that although you can get into university, you are then more likely to struggle and eventually 'drop out'. This would imply that the focus of widening participation policy needs to be on facilitating degree completion by poorer students, rather than just encouraging participation. Alternatively, is it simply the case that poorer students drop out of university not because they are poor but because they have lower levels of prior education achievement and are therefore less well prepared for college? It may be that poorer students are no more likely to drop out than other more advantaged students with similar levels of (low) prior achievement.

With our data we unfortunately cannot identify what specifically causes a student to drop out of university. For instance, we cannot differentiate between students who fail to meet the necessary academic standards and are forced to drop out of college because of this from students who simply choose to withdraw. Certainly different causes of drop out will have different policy solutions. Indeed there is a literature, which we discuss later in the paper, that acknowledges that some drop out may be rational if students drop out when they find the costs (psychic or monetary) exceed the potential benefits broadly defined (Manski (1989); Hartog et al 1989; Oosterbeek 1989; and Altonji 1993). Government intervention to prevent rational drop out would seem inadvisable. However, although we cannot identify the causes of drop out in our data, what we can do is identify the importance of pupils' prior academic achievement, as compared to their family background, in explaining drop out. This is important because if poor academic preparation is largely to blame for drop out, the policy solution would include improved school based preparation for university, as well as academic remedial help, rather than say financial assistance. This is particularly important for the English context as the socio-economic gap in HE

[2] The socio-economic gap in drop out from college in the US system is discussed extensively in Turner (2008).
[3] In this paper we use the terms college, university and higher education (HE) inter-changeably.

participation in the UK has indeed been found to be largely attributable to poor academic preparation by lower income students rather than barriers at the point of entry into university, such as credit constraints (Chowdry et al. 2008). As such, the policy solution to under *representation* of poorer students in higher education must be at least partially school focused and would need to raise the academic achievement of deprived children. It would be informative to determine whether drop out from university is also largely attributable to weak academic preparation as opposed to other family background related factors.

To address these questions, we use administrative data on an entire cohort[4] of young people in England who potentially could enter university in 2004/5 (at age 18). These data are unique in that they include information on each pupil's personal characteristics (e.g. ethnicity, date of birth and indicators of their socio-economic background) and also provide a complete record of each child's educational achievement from age 11 onwards[5]. This is the first time that such longitudinal data have been available to study the important issue of college drop out in the UK context.

The paper starts with a brief description of the English university system in section 2, to provide context to the paper. We then proceed to describe the previous literature (section 3) and the data (section 4). Section 5 gives the econometric model and section 6 the results. Section 7 concludes.

2. The English higher education system

For context, we briefly describe the English university system. In England, students who apply to university straight after secondary school (one fifth of the cohort) apply through a centralized admissions system called the University Centralised Admission System or UCAS. Students have to apply to specific degree courses i.e. they have to pre-select a specific degree subject in a specific institution. Hence by contrast to the

[4] We have data on all students in schools funded by the state in England who turned age 18 in 2003/4, i.e. more than half a million students. We exclude students in private schools. To the extent that such students have lower university drop out rates, we may be understating the socio-economic gap in university drop out.

[5] In the UK pupils take achievement tests or externally validated examinations at ages 11, 14, 16 and 18, hence we have a complete record of educational achievement from age 11 to age 19.

US system, the process of applying to university is more centrally coordinated and students are limited in the number of different institutions they can apply to (they can choose up to 5). It is usual for students to apply to study the same or similar degree subjects at a range of different institutions (although some students may only specify one or two options). Students can then be offered a place on a particular degree course subject to academic conditions, such as minimum grades achieved at age 18 (in their Advanced (A level) examinations which are equivalent to high school graduation). The results of students' examinations at age 18 are not generally known before application to university which means there may be a potential mismatch between students' expectations about their A level grades (which will determine their choice of degree subject and institution) and their actual A level grades.

In England it normally takes three years to achieve a Bachelors degree and, again unlike the US, transfer between institutions is not usual (fewer than 3% switch institutions). This is partly because there is no agreed system of college credits so transfer is difficult. Students' progression from one year to the next in their degree course is subject to a minimum academic level of achievement. This means that potentially students may drop out either because they fail to reach the required academic standard (the extent to which they have to pass examinations to progress from one year to the next varies by institution but all have some form of assessment) or because they anticipate that they will fail. Students are allowed to repeat parts of their degree that they have failed. If a student fails an element of their degree (a module) they may retake it the following year. However, there are limits (that vary by institution) to the number of times a student can retake an element of their degree course. Students that persistently fail will not be allowed to continue on their course.

As in the US, in England there are concerns that financial barriers may both prevent students from enrolling in HE in the first place and then cause them to drop out (Callender, 2003). However, the English system differs substantially from the US system in terms of the costs of study, with the former system being appreciably more generous, in terms of the state's contribution to the costs of university. English universities have historically been free to students, with costs paid by the general tax payer. This has changed in the last decade and institutions can now charge English

students up to £3000 per annum (although foreign students will pay more). These fees are means tested, with the poorest students fully exempt from fees. Repayment is also now on an income contingent basis after graduation, to minimize the disincentive effect on students from disadvantaged socio-economic backgrounds. Furthermore, students also have access to loans to fund their living costs during their degree.

One feature of the English university system that is particularly important for this paper is the fact that, unlike in the US and countries such as New Zealand and Sweden, England has historically had low levels of student 'drop out' from higher education (Dearing, 1997; NAO, 2007). Recent data suggests that 91.6% of full time students starting university in 2004/05 continued into their second year and 78.1% are expected to complete their degree (NAO, 2007)[6]. However, as the sector has expanded and the rate of non-completion has risen (Johnes and McNabb, 2004), so policy attention has shifted to this issue, and non-completion rates are now part of a range of indicators used by government to measure university performance. Indeed university league tables are produced in UK newspapers, ranking universities on a number of metrics, including their 'drop out' rate. These league tables generally do not control for student characteristics, and therefore may give a misleading impression of the true institutional quality in terms of retaining students, particularly for universities with a larger number of deprived and lower achieving students. Ideally value added models, which control for students' prior achievement, are needed to assess whether institutions have particularly high or low drop out rates relative to their student intake. Such models have not been possible previously, due to limited data. However, in this paper we show how such models may be operationalised using administrative data.

3. Previous literature

There is a large and growing literature on the role of family background, i.e. income and socio-economic status, in determining education outcomes, particularly university achievement (Blanden and Gregg, 2004; Carneiro and Heckman, 2002 and 2003; Gayle et al. 2002; Meghir and Palme, 2005; Haveman and Wolfe, 1993). Such studies

[6] The drop out rate in our data is somewhat lower than this as we focus specifically on young full time entrants. Comparable HESA data indicates that the UK non-continuation rate from the first year of a degree to the second for young full time entrants was 7.2% in 04/05. In our data for England the figure is very similar (just over 6%).

have found family background to be an important determinant of educational achievement and have also suggested that the socio-economic gap in educational achievement emerges early (see CMPO, 2006 and Feinstein, 2003 for the UK; Cunha and Heckman, 2007 and Cunha et al., 2006 for the US). In fact Cunha et al. (2006) concludes that family background and specifically credit constraints play only a limited role in determining HE participation, conditional on achievement in secondary school, although some recent studies dispute this (Belley and Lochner, 2007).

Even if education inequality emerges early, in the US at least, the raw socio-economic gap widens substantially if one measures Bachelor degree completion as opposed to enrolment (Turner, 2008). This raises the question as to whether the conditional drop out rate is higher for lower socio-economic group students, taking account of their prior education achievement. The literature on the relationship between socio-economic background and drop out from university is limited, although the US evidence finds differential drop out by family income (see Corrigan, 2003, Haveman and Wilson, 2005 and related issues in Bound, Lovenheim and Turner, 2008).

In the English context, Johnes & McNabb (2004) analyzed students entering and leaving the 'old' (pre-1992[7]) universities and distinguished between 'voluntary' drop out and 'involuntary' drop out i.e. failure. Jones & McNabb found that students from a lower socio-economic background were more likely to drop out voluntarily. Smith & Naylor (2001) used the same data to examine completion and non-completion. Using a binomial regression analysis of the probability that an individual withdraws from university for whatever reason, the authors found the risk of dropping out to be extremely high amongst students from lower social class backgrounds and living in high unemployment rate areas. More recently, Bennett (2003) showed self-declared financial hardship to be the most powerful predictor of a student's decision to withdraw from their degree course in the Business Studies department in a 'new' university in Greater London. Using data collected by the author himself, Bennett estimated a confirmatory factor analysis of the probability of non-completion among business students. He found self-declared financial hardship to be one of the strongest

[7] Prior to 1992, English higher education was divided into universities (higher status more academically oriented universities) and polytechnics (more vocationally oriented higher education institutions). In 1992 polytechnics became universities.

predictors of the individual quit decision. Other important factors also included low self-esteem and academic performance at university. Whilst these studies were able to control to some extent for a student's entry qualifications, they did not have rich data on students' prior achievement, and only considered a sub-set of UK universities.

This paper therefore estimates models of student drop out or non-completion using all universities in England. It is worth considering why we might be concerned with student drop out. Firstly, there may be economic costs associated with non-completion; there may have been a waste of resources if a student starts but does not complete a course (Yorke, 1998). Another potential concern is the sense of failure that a student may feel after dropping out of university and the impact on his or her productivity and earnings. That said we must be cautious in interpreting our results. Although the UK does not have the long standing US tradition of part time and intermittent higher education, increasingly UK students do enroll part time and progression may not be linear. This is particularly true of mature students (McGivney, 1996). Labelling (temporary) withdrawal as academic failure or wastage would seem inappropriate: just because students withdraw from their studies does not mean that they have not received any benefit from university (Johnes and Taylor, 1989). This is not merely a semantic debate: universities in England now face clear incentives to encourage student completion in the 'normal' time and non-completion whatever the cause is penalized financially[8]. If indeed poorer students are more likely to drop out than their more advantaged counterparts for a given level of prior achievement, this may lead to a tension between the widening participation agenda that encourages greater participation by deprived students and the desire by universities not to incur penalties from high levels of student withdrawal (Palmer, 2001).

Some economists have also made the argument that drop out from HE is efficient: weaker students who would not benefit from completing their degrees rightly drop out. Manski (1989) for example, argues that lower dropout rates would not necessarily make society better off. He suggests the decision to enrol is a decision to initiate an experiment, a possible outcome of which is dropout (see also Hartog et al 1989; Oosterbeek 1989; and Altonji 1993). Thus enrolment in HE incurs a risk for all

[8] Public funding received by each university in England is potentially affected by their performance. Non-completion is one measure of university performance used by the authorities.

students, namely the risk that they may have to drop out for whatever reason. Poorer students may face higher levels of this risk. For example, they may be more likely to fail to reach the level of educational achievement required or make their decisions about choice of institution and subject of study on the basis of poorer quality information. This higher level of risk may partially explain lower participation rates by poorer students. Even if poorer students face the same risk of drop out as their advantaged peers, if they are more risk adverse (Callender, 2003), then this too would at least partially explain their lower enrolment rate.

4. Data

We use linked administrative data sets to carry out the analysis: namely, the English National Pupil Database (NPD) / Pupil Level Annual School Census (PLASC) and individual student records maintained by the Higher Education Statistics Agency (HESA). The school administrative data (NPD/PLASC)[9] contain each pupil's record of their primary and secondary schooling, including information on their educational attainment from age 11 to 18, their personal characteristics, such as date of birth, ethnicity, home postcode, entitlement to free school meals[10] and whether English is an additional language in their home. The university records (HESA data) contain information on the degree subject, institution and other details of each student's university education, for all students studying for a first degree at UK universities. With these two sources of data linked together[11], we have longitudinal data on a cohort of students from age 11 through to potential HE participation at age 18 in 2004/5 and continued participation at age 19 in 2005/6. For the purposes of this paper, we consider only HE participants in English institutions and have a sample size of 121,827 observations from 120 HE institutions.

The dependent variable of interest is simply whether or not the pupil continued from one year to the next, i.e. continued in the same university from 04/05 to 05/06. We recognize that we are only able to consider drop out from the first year of study and if

[9] These data are maintained by the Department for Children, Schools and Families.
[10] This can be thought of as a proxy for very low family income. Pupils are eligible for Free School Meals (FSM) if their parents receive Income Support, income-based Jobseeker's Allowance, or Child Tax Credits, with a gross household income of less than £14,495 (in 2007–08 prices).
[11] School administrative records are fuzzy matched to higher education administrative records, using a variety of variables including name, date of birth and postcode. The matching process was carried out by the Department for Children, Schools and Families.

students from different socio-economic backgrounds tend to drop out at different stages, we may have partial results. For instance, if more advantaged students manage to remain in university longer, we may overstate the socio-economic gap in drop out in comparison to what we might expect to see in terms of degree completion. Around 6% of pupils failed to progress from one year to the next, indicating they dropped out from their institution (voluntarily or involuntarily) or decided to move to another institution in the following year. As our cohort only potentially entered HE the previous year, we are essentially measuring drop out after the first year of 3 years of study[12]. With our data we are not able to separate out those who drop out of higher education altogether, from those who switch institutions. Data from the Higher Education Statistics Agency suggest that 2.9% of students from low income neighborhoods switch institutions (across all years of enrolment) as compared to 2.7% of students from higher income neighbourhoods [13]. Since switching occurs at a similar rate across different types of student, this problem would not appear to be a major source of bias in our results.

A key feature of the data we use is that they include test score information on pupils from age 11 onwards. The test score information comes from age 11 (Key Stage 2) and age 14 (Key Stage 3) tests. These are national achievement tests sat by all children in state schools in England in English, Mathematics and Science. The tests are externally validated i.e. they are marked by individuals outside of the child's school. We include test scores in three subject areas – English, Mathematics and Science[14], using a non linear specification. The test data are supplemented by the results from public examinations taken by most students at age 16, namely General Certificates of Secondary Education (GCSEs), and for some students, Advanced levels (A-levels) at age 18. For GCSE, we use the capped total point score: this gives the total number of points accumulated from the student's eight highest GCSE grades.[15] At 18, we use the total (uncapped) point score. We divide the population into

[12] The vast majority of UK Bachelors degrees are 3 years of study.
[13] Higher Education Statistics Agency:
http://www.hesa.ac.uk/dox/performanceIndicators/0607/t3b_0607.xls
[14] We are grateful to an anonymous referee for guidance on the use of these scores.
[15] We use a capped total point score to avoid conflating the quantity of GCSEs taken and the grades received, as students may take a varying number of qualifications. For example, receiving 10 Grade D GCSEs would be equivalent (in terms of total points scored) to receiving 8 Grade C GCSEs (using the old tariff system), while we may not believe these are equivalent in terms of achievement.

five evenly sized quintile groups ranked according to their score at GCSE and A level or equivalent[16] to capture attainment at these levels. All in all these data contain the richest possible information on students' prior achievement to better enable us to identify the distinct role of academic preparation and socio-economic background in dropping out of university.

Based on the person's university, we also linked in an institution-level indicator of the university's research quality from the 2001 Research Assessment Exercise (RAE)[17]. We then combine this indicator of the quality of each institution's research, with an indicator of whether or not the institution is a member of the Russell Group of universities, a self-defined elite grouping of English universities. We define a high status institution as being all 20 of the research-intensive Russell Group institutions, plus any UK university with an average 2001 RAE rating that exceeds the lowest RAE score for a Russell Group university (see Chowdry et al. 2008 for further details and a list of institutions). In summary, we create a binary indicator of whether an institution is an elite institution or otherwise, recognizing the inherent problems in defining what constitutes quality in higher education from a student perspective.

The data do however, have a number of limitations. Firstly, the indicators for students' family background are problematic. Students who apply through the UCAS centralized application system described earlier are asked about the occupations of their parents. However, for students not applying through UCAS this information is absent altogether. For young full time entrants to a first degree, if we rely on data on parental socio-economic background, we have missing data for around one in five students. This is a long recognized problem in the UK. For instance, the UK National Audit Office have stressed the poor quality of the data available to identify the socio-economic background of students in HE.

An alternative to using parental SES is to use administrative records from earlier in these individuals' schooling. We do have an indicator of whether or not a student was eligible for free school meals (FSM) in secondary school. This indicator is limited by

[16] For students taking vocational qualifications at age 18 instead of A-levels, we have their point score using the official equivalencies between vocational and academic qualifications.
[17] The RAE is a quality assessment exercise to assess research quality across the HE sector in England and Wales. Quality assessment is based on peer review.

not being contemporaneous with HE participation and also by measuring the very bottom of the income distribution. Around 5% of students entering HE were eligible for free school meals in secondary school. We also have each pupil's postcode (similar to a zipcode) in secondary school when they were applying to enter university, and we can use this geographical marker[18] to link in information on the characteristics of the pupil's neighborhood. Specifically we can link in measures of socio-economic deprivation. (namely an Index of Multiple Deprivation (IMD) score, derived from Census data on the characteristics of individuals living in their neighborhood[19] and an Income Deprivation Affecting Children Index (IDACI) score, again constructed on the basis of Census data on individuals living in their neighborhood[20]). Whilst these administrative measures of family deprivation are not ideal (family income would be preferable, for example), we use individual and administrative measures of socio-economic background and test the robustness of our findings to using these alternative indicators. In our main table, we include individual measures of parental socio-economic status where available, with a missing dummy for those who do not provide this information. We then include our Free School Meal indicator and geographical measures which should capture, at least to some extent, the socio-economic circumstances of those pupils for whom we have missing individual level data.

Another limitation of the data is that we only consider young HE participants i.e. those participating at age 18. The drop out behavior of older HE participants may differ so our results are not necessarily generizable to older students. Finally, we only have data on state school pupils. A significant minority of students in England attend private schools prior to entering HE (just under 7% at age 16 in our data). If these more advantaged students were included in our sample and if they have very low drop

[18] See Geronimus, Bound and Neidert (1995) for a discussion of the limitations of using geographical based measures of socio-economic status in the context of health research. They conclude that there is a tendency for geographical measures to exaggerate the apparent effect of socio-economic characteristics on health outcomes. However, they acknowledge their findings are data specific and urge caution when making inference from geographically based measures. In this paper we explore both individual and geographical measures to address this problem.

[19] This is available at Super Output Area (SOA) level (comprising approximately 700 households), and makes use of information from seven different domains: income; employment; health and disability; education, skills and training; barriers to housing and services; living environment; and crime.

[20] IDACI is an additional supplementary element to the Index of Multiple Deprivation.

out rates, then our estimates of the socio-economic gap in HE drop out may well be lower bounds.

Table 1: Descriptive Statistics of Selected Variables

Variables	Non-participants	Enter HE at 18			Enter HE at 19
	All	All HE participants	Continued to year 2	Dropped out year 1	All
Social class: Professional occupations	-	0.204	0.208	0.116	0.197
		(0.403)	(0.406)	(0.320)	(0.397)
Social class: Associate professional and technical	-	0.118	0.119	0.101	0.114
		(0.323)	(0.324)	(0.302)	(0.318)
Social class: Administrative and secretarial	-	0.080	0.081	0.066	0.069
		(0.272)	(0.273)	(0.248)	(0.254)
Social class: Skilled trades	-	0.092	0.091	0.094	0.080
		(0.289)	(0.288)	(0.292)	(0.272)
Social class: Personal service	-	0.027	0.026	0.033	0.029
		(0.162)	(0.161)	(0.180)	(0.170)
Social class: Sales and customer service	-	0.021	0.021	0.029	0.027
		(0.145)	(0.144)	(0.168)	(0.163)
Social class: Process, plant and machine	-	0.062	0.062	0.061	0.054
		(0.241)	(0.242)	(0.239)	(0.226)
Social class: Elementary	-	0.036	0.036	0.037	0.037
		(0.187)	(0.187)	(0.189)	(0.190)
Social class: Missing	-	0.153	0.145	0.304	0.213
		(0.360)	(0.353)	(0.460)	(0.409)
Multiple deprivation index	0.216	-0.557	-0.569	-0.374	-0.397
	(1.517)	(1.157)	(1.149)	(1.254)	(1.284)
Multiple deprivation index squared	2.349	1.649	1.645	1.714	1.808
	(3.825)	(2.174)	(2.155)	(2.446)	(2.526)
Multiple deprivation index missing	0.122	0.126	0.128	0.099	0.136
	(0.328)	(0.332)	(0.334)	(0.298)	(0.343)
OA education index	0.222	0.297	0.299	0.273	0.307
	(0.121)	(0.128)	(0.128)	(0.129)	(0.140)
OA education index squared	0.064	0.105	0.105	0.091	0.114
	(0.074)	(0.090)	(0.090)	(0.087)	(0.105)
OA education index missing	0.122	0.126	0.128	0.098	0.136
	(0.327)	(0.332)	(0.334)	(0.298)	(0.343)
Male	0.536	0.432	0.432	0.429	0.477

	(0.498)	(0.495)	(0.495)	(0.495)	(0.499)
Free school meal	0.156	0.049	0.048	0.074	0.073
	(0.363)	(0.217)	(0.214)	(0.262)	(0.260)
Other White	0.022	0.025	0.025	0.025	0.029
	(0.148)	(0.157)	(0.157)	(0.156)	(0.169)
Black African	0.010	0.012	0.013	0.009	0.019
	(0.103)	(0.112)	(0.113)	(0.095)	(0.138)
Black Caribbean	0.013	0.009	0.009	0.013	0.013
	(0.116)	(0.097)	(0.095)	(0.114)	(0.116)
Other Black	0.006	0.004	0.004	0.003	0.006
	(0.083)	(0.066)	(0.066)	(0.058)	(0.080)
Indian	0.015	0.051	0.053	0.023	0.045
	(0.124)	(0.221)	(0.224)	(0.151)	(0.207)
Pakistani	0.022	0.025	0.025	0.022	0.036
	(0.147)	(0.156)	(0.156)	(0.147)	(0.186)
Bangladeshi	0.008	0.009	0.009	0.007	0.010
	(0.090)	(0.096)	(0.097)	(0.086)	(0.101)
Chinese	0.002	0.008	0.008	0.001	0.006
	(0.048)	(0.089)	(0.091)	(0.043)	(0.080)
Other Asian	0.001	0.005	0.005	0.002	0.004
	(0.034)	(0.074)	(0.075)	(0.053)	(0.068)
Mixed ethnicity	0.003	0.010	0.010	0.005	0.009
	(0.057)	(0.101)	(0.103)	(0.075)	(0.096)
Other ethnicity	0.014	0.014	0.014	0.012	0.018
	(0.119)	(0.118)	(0.118)	(0.112)	(0.134)
Unknown ethnicity	0.123	0.101	0.102	0.085	0.115
	(0.329)	(0.301)	(0.303)	(0.280)	(0.320)
Key stage 2: English (in level)	4.027	4.721	4.734	4.526	4.597
	(0.756)	(0.505)	(0.499)	(0.567)	(0.582)
Key stage 2: Maths (in level)	4.029	4.761	4.777	4.506	4.611
	(0.834)	(0.602)	(0.593)	(0.670)	(0.676)
Key stage 2: Science (in level)	4.174	4.753	4.764	4.575	4.649
	(0.711)	(0.523)	(0.518)	(0.568)	(0.586)
Key stage 3: English (in level)	5.034	6.157	6.179	5.827	5.934
	(1.068)	(0.793)	(0.784)	(0.848)	(0.875)
Key stage 3: Maths (in level)	5.248	6.716	6.750	6.183	6.400
	(1.158)	(0.974)	(0.959)	(1.042)	(1.087)
Key stage 3: Science (in level)	4.948	6.178	6.206	5.750	5.914
	(1.023)	(0.850)	(0.841)	(0.876)	(0.950)
Age 16 examination scores General Certificates of Secondary Education (capped) proportion in highest quintile	0.068	0.511	0.528	0.239	0.376
	(0.252)	(0.499)	(0.499)	(0.426)	(0.484)
Age 18 examination scores	24.335	288.365	295.875	169.512	193.977

Advanced level or equiv point score					
	(72.898)	(144.308)	(141.201)	(140.881)	(166.417)
Level 3 at 18	0.202	0.952	0.962	0.786	0.921
	(0.402)	(0.212)	(0.188)	(0.409)	(0.268)
Subject: Allied to medicine	-	0.071	0.072	0.067	-
		(0.258)	(0.258)	(0.251)	
Subject: Biological science	-	0.106	0.108	0.067	-
		(0.307)	(0.311)	(0.251)	
Subject: Veterinary science	-	0.005	0.005	0.006	-
		(0.072)	(0.071)	(0.078)	
Subject: Agriculture	-	0.006	0.005	0.016	-
		(0.080)	(0.077)	(0.126)	
Subject: Physics	-	0.048	0.049	0.022	-
		(0.214)	(0.217)	(0.147)	
Subject: Mathematics	-	0.025	0.026	0.008	-
		(0.157)	(0.161)	(0.090)	
Subject: Computer science	-	0.050	0.050	0.052	-
		(0.218)	(0.218)	(0.223)	
Subject: Engineer & Technology	-	0.055	0.055	0.045	-
		(0.228)	(0.229)	(0.209)	
Subject: Architecture	-	0.018	0.018	0.013	-
		(0.134)	(0.135)	(0.117)	
Subject: Social studies	-	0.087	0.089	0.059	-
		(0.282)	(0.285)	(0.236)	
Subject: Law	-	0.052	0.054	0.026	-
		(0.223)	(0.227)	(0.161)	
Subject: Business	-	0.104	0.105	0.099	-
		(0.306)	(0.306)	(0.298)	
Subject: Mass communication	-	0.036	0.036	0.041	-
		(0.188)	(0.187)	(0.199)	
Subject: Languages	-	0.103	0.106	0.057	-
		(0.304)	(0.308)	(0.232)	
Subject: History	-	0.021	0.022	0.010	-
		(0.145)	(0.148)	(0.099)	
Subject: Arts	-	0.095	0.091	0.155	-
		(0.293)	(0.288)	(0.362)	
Subject: Education	-	0.033	0.033	0.032	-
		(0.180)	(0.180)	(0.177)	
Subject: Combined	-	0.028	0.028	0.035	-
		(0.167)	(0.166)	(0.185)	-
University status index	-	0.374	0.387	0.170	0.279
		(0.484)	(0.487)	(0.375)	(0.448)
N	435596	121827	114586	7241	56001

Data source: linked National Pupil Database/ Pupil Level Annual School Census/ Individual Learner Record and Higher Education Statistics Agency data. Mean proportions given with standard errors in parentheses.

In Table 1, we show full descriptive statistics for our sample, including a comparison with students who do not enroll in university (column 1) and indeed those who entered HE one year later at age 19 (final column). Non participants are more likely to be male and white. What is most striking however, is that whilst only 6% of the sample of non participants achieve in the top quintile of GCSE achievement (at age 16), around half of those enrolling in higher education have this higher level of achievement. At age 18, again the non-participants have much lower levels of achievement, primarily because most of them will have left the education system by that age and would not have attempted to take A levels or high school graduation equivalent qualifications (thereby getting a zero score). Non-participants are also more likely to be deprived, as measured by both their free school meal status and their geographical based measure of deprivation.

Moving from left to right across the table, we now consider the sample who enrolled in university at age 18. From this sample, 'drop outs' are slightly (but not significantly) more likely to be white. Those who do not continue beyond their first year of study are lower achievers at age 16 (GCSE) and at age 18 (A level). They are less likely to have a parent in a professional occupation and are slightly more deprived on our measure of deprivation. They are much less likely to attend a high status university.

Whilst we cannot include those who enter HE at 19 in 2005/6 in our analysis (because we do not have data on these students the following year to measure their propensity to withdraw), it is useful to see how the characteristics of later entrants compare with our sample. In general, those entering later appear to be more likely to be non-white, have lower achievement, are somewhat more deprived and have a lower probability of attending a high status institution. Such differences are not statistically significant however.

5. Econometric model

We define a model of non-continuation as follows:

$$\Pr(Y = 1 \mid X = x) = \Phi(x'\beta),$$
$$Y_{ij} = X_{ij}'\beta + \varepsilon_{ij}, \qquad (1)$$

where Y takes a value of 1 if individual i from university j withdraws from the institution in the following academic year. We cannot distinguish between voluntary and involuntary withdrawal.

X_{ij}' is a vector of student personal characteristics taken largely from the secondary school administrative data set. This includes date of birth, gender, ethnicity, disability status, and English as an Additional Language. Our explanatory variables of interest are firstly our measures of the student's socio-economic status, measured in a number of ways, as described in the previous section. Our second set of explanatory variables of interest are our comprehensive measures of prior educational attainment, namely test scores at age 11, 14, 16 (GCSE) and age 18 (A-level or high school graduation equivalent). The parameter ε_{ij} denotes the error term. The model is then estimated using a probit model, and marginal effects are reported in all tables, standard errors are clustered by university. We also allow for the potentially heterogeneous effects from socio-economic status by exploring various interactions as discussed in the section below.

In some specifications we also include a vector of variables describing the nature of the individual's university and degree subject. We also include detailed information on university quality, based on the institution-level Research Assessment Exercise (RAE) scores for 2001 and a dummy variable representing whether the institution is an elite Russell group university. When we investigate interactions, we simplify this information into a simple binary variable indicating whether the university attended is a high status institution, as described earlier in the data section.

There are a number of estimation issues. We are attempting to determine whether or not poorer students are more likely to drop out of university, allowing for their prior educational achievement. Whilst we are confident that we can control fully for each pupil's prior achievement, which previous research has been unable to do, we recognize that there may be unobserved factors that determine whether or not a pupil

drops out of university and that these may be correlated with their socio-economic status. For example, poorer students may tend to enroll in certain types of universities. There may be unobserved features of a student's university that then determine their decision to withdraw. Whilst we control for degree subject and institution quality in a relatively comprehensive way, there are many other features of universities that we are unable to allow for, such as pastoral care, advice and guidance given etc. To partially address this we therefore estimate models which allow for mean differences in drop out across different institutions. We do this by including a dummy variable for each university, which allows for any differences in the non-continuation rate across different universities that are not controlled for by our explanatory variables. The variable λ_j represents institutional dummy variables, which can be incorporated to give the following equation:

$$\Pr(Y = 1 \mid X = x) = \Phi(x'\beta),$$
$$Y_{ij} = X_{ij}'\beta + \lambda_{j+}\varepsilon_{ij},$$

At the suggestion of an anonymous referee, we estimated an HE participation equation for everyone in our sample. We then obtained a predicted likelihood of HE participation and included this predicted probability in the drop out equation. The variable measuring the individual's predicted probability of HE participation was significantly negatively correlated with whether or not the person dropped out of university. In other words, those more likely to participate in HE were certainly more likely to drop out. Inclusion of this predicted probability in the drop out equation did not qualitatively alter our key results. We acknowledge however, that we do not have an identification strategy that enables us to separately identify the enrollment and the drop out decisions[21].

We recognize that we have not fully overcome the potential endogeneity of university participation and choice of institution. Our results are therefore descriptive rather than necessarily causal. However, we are confident that we are able to add to the literature in a meaningful way as, for the first time using English data, we are able to properly control for pupils' entire educational trajectory from the age of 11 when modeling their tendency to drop out of university.

[21] Results available on request.

6. Results

Table 2 below estimates the model described above, where the dependent variable takes a value of one if the person continued beyond their first year of study in the same university and zero otherwise. We then add sequentially various explanatory variables as controls. Column 1 controls for the student's socio-economic background and individual characteristics only. The probability of 'dropping out' varies by socio-economic background. In broad terms, students from higher socio-economic status backgrounds, who live in less materially deprived and more educated neighborhoods, have lower rates of drop out. It is noticeable that the geographical based measures remain significant even with the inclusion of individual level measures of socio-economic background. The fifth of the sample who do not have an individual level indicator of socio-economic background (i.e. who have a missing dummy variable for socio-economic status) are significantly (by 9 percentage points) more likely to drop out of university than pupils from a managerial/ senior official background. For these students with missing data, the only indicators of socio-economic background that we have are the geographical measures.

To illustrate the magnitude of socio-economic differences in drop out, a pupil from a professional background is 1.3 percentage points less likely to drop out than a student from a managerial/senior official background. Equally a pupil whose parents are in sales and customer service occupations is around 3 percentage points more likely to drop out than a student from a managerial/senior official background. In the context of low overall drop our rates (6% in our sample), these socio-economic differences are large.

In terms of other individual characteristics, there also appears to be an ethnicity dimension to the problem of drop out. Almost all groups of ethnic minority students are significantly *less* likely to dropout compared to white students. This finding is consistent with a number of recent UK studies that have found that ethnic minority students make more progress academically, particularly in secondary school, than Whites (Wilson, Burgess and Briggs, 2005). Indeed many ethnic minority groups are also more likely to attend college in the UK even without allowing for prior

educational achievement. For a given level of prior achievement in secondary school, all ethnic minority groups are significantly more likely to enroll in college as compared to Whites (Chowdry et al. 2008). In our data, Black African, Pakistani and Bangladeshi students are approximately 3 percentage-points less likely to withdraw than Whites. Chinese students are around 4 percentage-points less likely to dropout compared to students of white ethnic background. Other individual characteristics, such as gender, are not significant.

Column 3 then adds controls for prior achievement. Specifically we include age 11 and 14 test scores, as well as GCSE and A level scores and a dummy variable indicating whether or not the individual achieved an A level or high school graduation equivalent qualification (level 3)[22]. Whilst the age 11 test scores (Key Stage 2) are generally insignificant, some of the age 14 scores in mathematics (Key Stage 3) are significantly correlated with drop out. What is striking however, is the strong relationship between prior achievement at age 16 (GCSE) and 18 (A level) with drop out. When we include these prior attainment variables the coefficients on the socio-economic variables are reduced, though most of the socio-economic coefficients remain statistically significant. For instance, students from a professional background are now 0.6 percentage points less likely to drop out than students from a managerial/ senior official background. Students with a parent in a sales or customer service occupation are now 2 percentage points more likely to drop out than students from a managerial/ senior official background. This result implies that not all the socio-economic gradient in 'drop out' is attributable to the poorer prior academic achievement of disadvantaged pupils.

In the light of previous comments about sorting across HE institutions by socio-economic background, in column 4 we control for degree subject and institution status. The institution status and degree subject variables are insignificant[23] and there

[22] In other words we control for whether the student reached a high school graduate level of achievement via an alternative (normally vocational) qualification.

[23] Note that in models without measures of prior achievement, there are large statistically significant differences in the likelihood of drop out by degree subject and indeed by institution status. However, inclusion of the prior achievement variables renders these variables insignificant. Some of the conditional differences in the likelihood of drop out across degree subject, controlling only for socio-economic background and individual characteristics, are very large indeed (e.g. agricultural students were 18 percentage points more likely to drop out than medical students). This may be of policy interest but is not the focus of this paper.

are only marginal reductions in the coefficients on the socio-economic variables. As discussed earlier, we also want to control for the unobserved characteristics of universities that may impact on student 'drop out'. We do this by including institution dummy indicator variables in Column 5. The inclusion of these dummy variables does not change the coefficients on the socio-economic variables in a major way: i.e. there remains a socio-economic gradient in university 'drop out' even after controlling for prior attainment and institution means. This suggests that the higher withdrawal rate of poorer students is not by and large attributable to the types of degree subjects they are studying and the institution type they enroll in. In any case, since students' choice of degree and institution is itself related to their socio-economic background, this specification removes some of the socio-economic effect we are trying to measure.

Table 2: The socio-economic gradient in HE non-continuation

	Socio-economic background and personal characteristics	**Plus prior attainment**	**Plus HE characteristics**	**Plus institution dummy variables**
Social class: Professional occupations	-0.0129	-0.0064	-0.0061	-0.0057
	[0.0021]**	[0.0019]**	[0.0018]**	[0.0017]**
Social class: Associate professional and technical occupations	0.0034	0.0017	0.0016	0.0015
	[0.0024]	[0.0022]	[0.0021]	[0.0021]
Social class: Administrative and secretarial occupations	0.0017	0.0015	0.0007	0.0002
	[0.0029]	[0.0025]	[0.0024]	[0.0023]
Social class: Skilled trades occupations	0.0119	0.0057	0.0045	0.0040
	[0.0032]**	[0.0025]*	[0.0024]	[0.0023]
Social class: Personal service occupations	0.0204	0.0080	0.0055	0.0057
	[0.0060]**	[0.0045]	[0.0039]	[0.0038]
Social class: Sales and customer service	0.0307	0.0166	0.0141	0.0128
	[0.0069]**	[0.0050]**	[0.0047]**	[0.0046]**
Social class: Process, plant and machine	0.0113	0.0061	0.0052	0.0050
	[0.0041]**	[0.0036]	[0.0034]	[0.0034]
Social class: Elementary occupation	0.0127	0.0030	0.0011	0.0005
	[0.0052]*	[0.0040]	[0.0037]	[0.0035]
Social class: Missing	0.0928	0.0403	0.0243	0.0221
	[0.0174]**	[0.0083]**	[0.0057]**	[0.0038]**
Multiple deprivation index	0.0043	0.0012	0.0025	0.0023

	[0.0017]*	[0.0013]	[0.0009]**	[0.0007]**
Multiple deprivation index squared	-0.0012	-0.0005	-0.0007	-0.0005
	[0.0004]**	[0.0004]	[0.0003]*	[0.0003]
Multiple deprivation index missing	0.0183	-0.0037	0.0013	0.0015
	[0.0325]	[0.0199]	[0.0220]	[0.0200]
OA education index	-0.1061	-0.0585	-0.0676	-0.0571
	[0.0227]**	[0.0192]**	[0.0180]**	[0.0162]**
OA education index squared	0.1021	0.0732	0.0713	0.0498
	[0.0304]**	[0.0259]**	[0.0245]**	[0.0214]*
OA education index missing	-0.0713	-0.0477	-0.0485	-0.0407
	[0.0103]**	[0.0110]**	[0.0094]**	[0.0099]**
Free school meal	0.0009	-0.0006	0.0008	0.0013
	[0.0029]	[0.0028]	[0.0026]	[0.0024]
Other White	-0.0078	-0.0080	-0.0075	-0.0080
	[0.0047]	[0.0041]	[0.0038]*	[0.0036]*
Black African	-0.0282	-0.0289	-0.0258	-0.0263
	[0.0050]**	[0.0032]**	[0.0030]**	[0.0023]**
Black Caribbean	-0.0014	-0.0186	-0.0198	-0.0210
	[0.0116]	[0.0046]**	[0.0036]**	[0.0029]**
Other Black	-0.0219	-0.0262	-0.0255	-0.0251
	[0.0068]**	[0.0043]**	[0.0041]**	[0.0037]**
Indian	-0.0346	-0.0311	-0.0285	-0.0265
	[0.0034]**	[0.0024]**	[0.0020]**	[0.0014]**
Pakistani	-0.0259	-0.0255	-0.0221	-0.0207
	[0.0036]**	[0.0025]**	[0.0024]**	[0.0020]**
Bangladeshi	-0.0317	-0.0269	-0.0234	-0.0211
	[0.0049]**	[0.0038]**	[0.0035]**	[0.0030]**
Chinese	-0.0429	-0.0349	-0.0332	-0.0322
	[0.0040]**	[0.0028]**	[0.0026]**	[0.0020]**
Other Asian	-0.0272	-0.0198	-0.0174	-0.0172
	[0.0058]**	[0.0054]**	[0.0055]**	[0.0051]**
Mixed ethnicity	-0.0270	-0.0199	-0.0197	-0.0188
	[0.0051]**	[0.0045]**	[0.0043]**	[0.0040]**
Other ethnicity	-0.0172	-0.0207	-0.0195	-0.0201
	[0.0051]**	[0.0032]**	[0.0029]**	[0.0024]**
Unknown ethnicity	-0.0016	-0.0016	-0.0014	-0.0012
	[0.0033]	[0.0027]	[0.0026]	[0.0026]
Male	0.0027	-0.0015	-0.0021	-0.0012
	[0.0024]	[0.0016]	[0.0015]	[0.0014]
Date of birth	0.0000	0.0000	0.0000	0.0000
	[0.0000]	[0.0000]*	[0.0000]*	[0.0000]**
Disabled	0.0082	-0.0041	-0.0060	-0.0081
	[0.0054]	[0.0037]	[0.0031]	[0.0021]**
Key stage 2 English: Level 2-2.99		0.0013	0.0027	0.0100

	[0.0228]	[0.0232]	[0.0257]
Key stage 2 English: Level 3-3.99	-0.0057	-0.0020	0.0019
	[0.0130]	[0.0138]	[0.0146]
Key stage 2 English: Level 4-4.99	-0.0001	0.0044	0.0079
	[0.0140]	[0.0143]	[0.0142]
Key stage 2 English: Level 5 and over	0.0054	0.0092	0.0136
	[0.0151]	[0.0156]	[0.0160]
Key stage 2 English: Missing	0.0080	0.0128	0.0164
	[0.0162]	[0.0177]	[0.0184]
Key stage 2 Maths: Level 2-2.99	-0.0023	0.0042	0.0029
	[0.0165]	[0.0185]	[0.0176]
Key stage 2 Maths: Level 3-3.99	0.0040	0.0070	0.0085
	[0.0112]	[0.0117]	[0.0111]
Key stage 2 Maths: Level 4-4.99	0.0022	0.0062	0.0073
	[0.0110]	[0.0109]	[0.0101]
Key stage 2 Maths: Level 5 and over	-0.0016	0.0038	0.0050
	[0.0111]	[0.0113]	[0.0104]
Key stage 2 Maths: Missing	-0.0068	-0.0035	-0.0021
	[0.0100]	[0.0109]	[0.0102]
Key stage 2 Science: Level 2-2.99	-0.0247	-0.0179	-0.0174
	[0.0109]*	[0.0132]	[0.0138]
Key stage 2 Science: Level 3-3.99	-0.0124	-0.0091	-0.0070
	[0.0114]	[0.0126]	[0.0138]
Key stage 2 Science: Level 4-4.99	-0.0113	-0.0078	-0.0054
	[0.0145]	[0.0149]	[0.0157]
Key stage 2 Science: Level 5 and over	-0.0100	-0.0065	-0.0039
	[0.0139]	[0.0144]	[0.0154]
Key stage 2 Science: Missing	-0.0109	-0.0067	-0.0040
	[0.0121]	[0.0132]	[0.0148]
Key stage 3 English: Level 2-2.99	0.0279	0.0261	0.0373
	[0.0390]	[0.0382]	[0.0435]
Key stage 3 English: Level 3-3.99	-0.0100	-0.0155	-0.0135
	[0.0201]	[0.0179]	[0.0188]
Key stage 3 English: Level 4-4.99	-0.0147	-0.0176	-0.0136
	[0.0184]	[0.0174]	[0.0192]
Key stage 3 English: Level 5 and over	-0.0131	-0.0181	-0.0116
	[0.0295]	[0.0334]	[0.0312]
Key stage 3 English: Missing	-0.0145	-0.0191	-0.0149
	[0.0197]	[0.0181]	[0.0195]
Key stage 3 Maths: Level 2-2.99	-0.0382	-0.0385	-0.0368
	[0.0048]**	[0.0022]**	[0.0012]**
Key stage 3 Maths: Level 3-3.99	-0.0361	-0.0357	-0.0342
	[0.0083]**	[0.0062]**	[0.0058]**
Key stage 3 Maths: Level 4-4.99	-0.0411	-0.0404	-0.0386

	[0.0074]**	[0.0059]**	[0.0055]**
Key stage 3 Maths: Level 5 and over	-0.1693	-0.1819	-0.1819
	[0.1199]	[0.1245]	[0.1317]
Key stage 3 Maths: Missing	-0.0460	-0.0452	-0.0433
	[0.0106]**	[0.0093]**	[0.0091]**
Key stage 3 Science: Level 2-2.99	0.0454	0.0635	0.0444
	[0.1085]	[0.1121]	[0.0955]
Key stage 3 Science: Level 3-3.99	0.0362	0.0323	0.0232
	[0.1007]	[0.0924]	[0.0828]
Key stage 3 Science: Level 4-4.99	0.0476	0.0470	0.0373
	[0.1062]	[0.1008]	[0.0921]
Key stage 3 Science: Level 5 and over	0.0277	0.0272	0.0233
	[0.0395]	[0.0363]	[0.0372]
Key stage 3 Science: Missing	0.0451	0.0476	0.0390
	[0.0995]	[0.0965]	[0.0896]
KS4 missing	-0.0175	-0.0123	-0.0120
	[0.0074]*	[0.0083]	[0.0077]
2nd capped KS4 quintile	-0.0164	-0.0114	-0.0116
	[0.0051]**	[0.0057]*	[0.0054]*
3rd capped KS4 quintile	-0.0388	-0.0289	-0.0268
	[0.0036]**	[0.0040]**	[0.0041]**
4th capped KS4 quintile	-0.0520	-0.0387	-0.0363
	[0.0054]**	[0.0051]**	[0.0051]**
Highest capped KS4 quintile	-0.0816	-0.0586	-0.0538
	[0.0099]**	[0.0083]**	[0.0082]**
KS5 missing	-0.0132	-0.0171	-0.0147
	[0.0036]**	[0.0033]**	[0.0023]**
KS5 points	-0.0002	-0.0002	-0.0002
	[0.0000]**	[0.0000]**	[0.0000]**
Level 3 at 18	-0.0322	-0.0228	-0.0195
	[0.0080]**	[0.0043]**	[0.0038]**
Level 3 at 18: Missing	-0.0144	-0.0048	0.0033
	[0.0142]	[0.0176]	[0.0199]
Subject: Missing		0.1909	0.1559
		[0.0383]**	[0.0338]**
Subject: Allied to medicine		-0.0088	-0.0081
		[0.0084]	[0.0082]
Subject: Biological science		-0.0016	-0.0035
		[0.0097]	[0.0090]
Subject: Veterinary science		0.0119	-0.0030
		[0.0173]	[0.0128]
Subject: Agriculture		0.0423	0.0097
		[0.0243]	[0.0139]
Subject: Physics		-0.0033	-0.0029

			[0.0083]	[0.0081]
Subject: Mathematics			-0.0023	0.0000
			[0.0082]	[0.0083]
Subject: Computer science			0.0061	0.0028
			[0.0111]	[0.0102]
Subject: Engineer & Technology			0.0062	0.0057
			[0.0113]	[0.0110]
Subject: Architecture			-0.0017	-0.0040
			[0.0101]	[0.0091]
Subject: Social studies			0.0040	0.0015
			[0.0101]	[0.0093]
Subject: Law			-0.0011	-0.0037
			[0.0094]	[0.0086]
Subject: Business			0.0097	0.0041
			[0.0110]	[0.0098]
Subject: Mass communication			0.0163	0.0037
			[0.0169]	[0.0103]
Subject: Languages			0.0041	0.0020
			[0.0099]	[0.0094]
Subject: History			-0.0024	-0.0031
			[0.0107]	[0.0102]
Subject: Arts			0.0293	0.0157
			[0.0166]	[0.0121]
Subject: Education			0.0005	-0.0051
			[0.0106]	[0.0088]
Subject: Combined			0.0104	0.0110
			[0.0128]	[0.0126]
Good university			0.0037	
			[0.0065]	
Institutional indicator variables	No	No	No	Yes
Observations	121827	121827	121827	121591
Log likelihood	-25683.5	-23825	-23071	-22577.9
Pseudo R-squared	0.0647	0.1324	0.1599	0.1774

Data source: linked National Pupil Database/ Pupil Level Annual School Census/ Individual Learner Record and Higher Education Statistics Agency data. This model is estimated using a probit model and the results presented are marginal effects. Standard errors are given in parentheses.
* indicates statistically significant at the 5% level. ** indicates statistically significant at the 1% level.
Base case is: managerial/ senior official social class, did not receive free school meals, white, female, not disabled, key stage 2 English score less than 2, key stage 2 maths score less than 2, key stage 2 science score less than 2, key stage 3 English score less than 2, key stage 3 maths score less than 2, key stage 3 science score less than 2, in lowest key stage 4 quintile, no level 3 qualification achieved at 18, base degree subject: medicine, attending institution not classified as one of the 'good' universities. Institution dummy variables are statistically significant at the 1% level.

Whilst we control for prior achievement in the models above, we also explored interactions between prior achievement and non-continuation. We therefore re-estimated the model separately for pupils in the top three quintiles of the age 11 test score distribution, as shown in Table 3. The general finding from Table 2 holds: i.e. there remain socio-economic differences in the tendency for pupils to drop out even after controlling fully for prior achievement. However, the individual level socio-economic background variables are by and large less significant than in the full model. For example, students who have a parent in a professional occupation and who achieve in the top quintile of age 11 achievement are around 5 percentage points less likely to drop out than pupils in the same quintile of age 11 achievement who have a parent in a managerial occupation. This result is significant. When we consider pupils in the third quintile of age 11 achievement (column 1) and the second quintile (column 2), we see a similar pattern. Pupils with a parent in a professional occupation are less likely to drop out than pupils with a parent in a managerial occupation. In fact the magnitude of the coefficient is if anything greater for pupils in the third quintile of age 11 achievement. However, the coefficients are insignificant and the standard errors large. This is of course partially driven by the smaller numbers of pupils with professional parents who achieve in the third quintile of age 11 achievement. By contrast, pupils in the third quintile of age 11 achievement who have a parent in a sales or customer service occupation are 3 percentage points more likely to drop out than pupils in that quintile of achievement with managerial parents. This result is significant but does not hold for pupils in the top quintiles of age 11 achievement.

Table 3: socio-economic gradient in HE non-continuation by Key Stage 2 achievement

	KS2: 2nd Quintile	KS2: 3rd Quintile	KS2: Top quintile
Social class: Professional occupations	-0.0085	-0.0034	-0.0045
	[0.0062]	[0.0042]	[0.0018]*
Social class: Associate professional and technical occupations	-0.0008	0.0068	-0.0002
	[0.0070]	[0.0041]	[0.0025]
Social class: Administrative and secretarial occupations	0.0043	-0.0029	-0.0015
	[0.0063]	[0.0044]	[0.0024]
Social class: Skilled trades occupations	0.0092	0.0042	0.0007
	[0.0073]	[0.0050]	[0.0030]
Social class: Personal service occupations	0.0135	0.0187	-0.0077
	[0.0115]	[0.0087]*	[0.0034]*
Social class: Sales and customer service	0.0338	0.0139	0.0007
	[0.0163]*	[0.0109]	[0.0044]
Social class: Process, plant and machine	-0.0065	0.0117	0.0033
	[0.0072]	[0.0061]	[0.0046]
Social class: Elementary occupation	0.0156	-0.0069	-0.0018
	[0.0093]	[0.0056]	[0.0045]
Social class: Missing	0.0305	0.0218	0.0164
	[0.0074]**	[0.0056]**	[0.0035]**
Multiple deprivation index	0.0023	0.0042	0.0011
	[0.0021]	[0.0016]**	[0.0010]
Multiple deprivation index squared	-0.0003	0.0000	-0.0007
	[0.0008]	[0.0006]	[0.0004]
Multiple deprivation index missing	-0.1068	0.0291	0.0140
	[0.0134]**	[0.0788]	[0.0419]
OA education index	-0.0181	-0.0682	-0.0569
	[0.0466]	[0.0384]	[0.0241]*
OA education index squared	-0.0378	0.0668	0.0610
	[0.0650]	[0.0512]	[0.0303]*
OA education index missing	0.9775	-0.0541	-0.0359
	[0.0002]**	[0.0125]**	[0.0111]**
Free school meal	0.0054	0.0019	0.0019
	[0.0078]	[0.0058]	[0.0044]
Observations	18411	29384	45882
Log likelihood	-4100	-5778	-6178
Pseudo R-squared	0.1657	0.1519	0.1558

Data source: linked National Pupil Database/ Pupil Level Annual School Census/ Individual Learner Record and Higher Education Statistics Agency data. This model is estimated using a probit model and the results presented are marginal effects. Standard errors are given in parentheses.

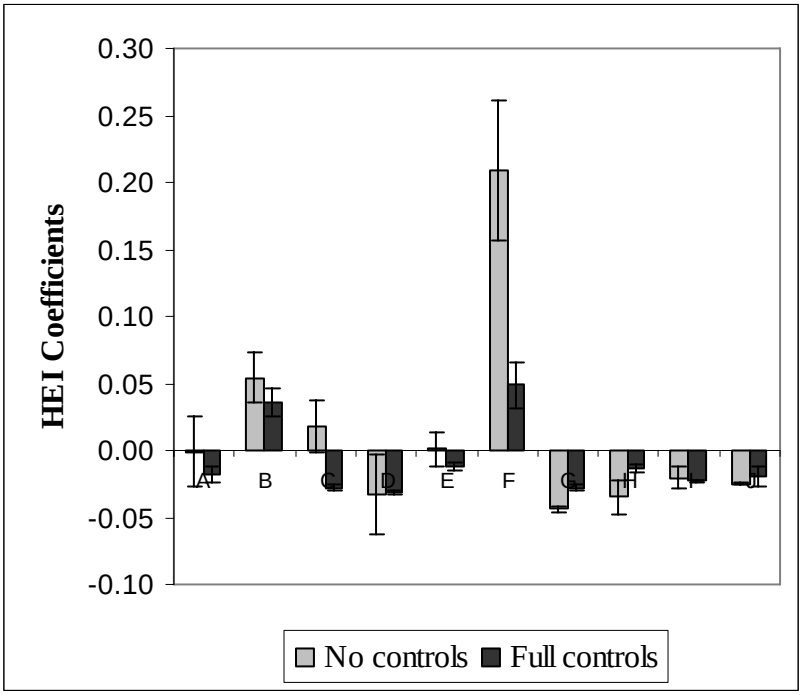
* indicates statistically significant at the 5% level. ** indicates statistically significant at the 1% level. Base case and controls are as in Column 4 of Table 2, which includes university indicator dummy variables.

Further interactions by gender and by university status and ethnicity were explored and reported in the appendix. It is difficult to discern clear patterns, not least because some (but not all) of the socio-economic background variables become insignificant when estimated separately by sub group. Whilst the coefficients on some variables do vary across these different groups of students and institutions, the results suggest that the relationship between socio-economic background and student 'drop out' is quite similar by gender and institution type (with a marginally smaller socio-economic gradient for students at higher status institutions and a higher socio-economic gradient for white students – but not significantly so). For Black students by contrast, there appears to be very little socio-economic gradient in 'drop out' once you control fully for prior achievement.

One motivation for this paper was to explore whether universities themselves have different rates of drop out, once we control for the prior achievement of their students. This is important given the policy importance of university 'league tables' in the UK and the notion of drop out rates being one metric by which university quality is judged. The model in column 4 of Table 2 allows for unobserved characteristics of universities that remain constant. The institutional dummy variables that are included measure the extent to which their drop out varies from the mean once one controls for student intake characteristics. Figure 1 investigates the magnitude of these institutional effects further. It shows the size and significance of 10 randomly chosen institutional effects from the model estimated in Table 2. In most cases shown, the magnitude of the coefficient on the institutional dummy variable without controls is not statistically significantly different from the magnitude when full controls for student prior achievement are included. Nonetheless for some specific institutions i.e. institution F, the inclusion of student characteristics and prior achievement makes a significant difference to the magnitude of the coefficient. In other words, for some institutions it may be that raw drop out rates can be misleading and that full account needs to be taken of their student intake. Of course one could argue that we should not be controlling for some student characteristics, e.g. ethnicity, since we do not want to argue that drop out rates vary inherently by this particular individual characteristic.

That said, it is useful to potential students and universities themselves to consider value added adjustments to drop out rates.

Figure 1: Randomly selected institutional coefficients with and without controls



F statistic for joint significance of institution dummy indicator variables (with full controls): χ^2_{91}	4.3e+07 [0.000]
Total number of institution coefficients per equation	119
N of institutions that we can reject the equality test at the 5% level, i.e. t>=1.96	67
% of institutions that experienced significant changes after control variables were added	56%

Note: the t-statistics are calculated using the following cross-equation formula: $\frac{\beta_1-\beta_2}{\sqrt{(\sigma_1^2+\sigma_2^2)}}$ (see Ferrer-i-Carbonell, 2005).
We used 4 standard-error bars (2 above and 2 below) to indicate 95% C.I.

7. Conclusions

Our results suggest that, in the UK, as in the US, there is a significant gap in the drop out rate between advantaged and disadvantaged pupils. Much of this gap disappears once we allow for students' prior achievement, suggesting that some of the apparent difference in first year drop out rates between richer and poorer students is actually attributable to differences in their academic preparation for HE and/or their ability, as measured by earlier measures of educational achievement. In the context, of a

relatively low aggregate rate of first year drop out (6%) from English universities, once we fully control for prior achievement, a pupil from a professional background is 0.6 percentage points less likely to drop out than a student from a managerial/senior official background. Equally a pupil whose parents are in sales and customer service occupations is 1.6 percentage points more likely to drop out than a student from a managerial/senior official background. In the context of a low overall drop out rate in the UK, these differences in drop out rates are arguably sizeable.

In this paper we were constrained by limitations with our data in terms of the quality of information on the socio-economic background of pupils. This is a real problem with UK administrative education data which needs to be resolved, perhaps by including information on parental education in administrative data sets. We overcame the difficulties by using multiple measures of socio-economic background, including individual measures of SES and geographical markers of neighborhood deprivation. Whilst this makes us more confident that we are controlling for socio-economic background adequately, it does mean that interpretation is problematic. Certainly when exploring interactions between socio-economic background and individual characteristics such as gender or ethnicity it is hard to discern clear patterns amongst the numerous socio-economic background indicators that we use. That said, it is clear from our results that significant differences in drop out remain across the different measures of socio-economic background, even after controlling for pupil characteristics and prior achievement. This is important from a policy perspective.

We also found that raw indicators of the drop out rate of English universities could be misleading for some institutions, if one's purpose is to use such measures as indicators of university efficiency. We conclude that the magnitude of universities' drop out rates can indeed change markedly in individual cases, if the prior achievement of students is taken fully into account. In policy terms, this suggests that if we are to use drop out rates as measures of institution performance, we must be careful to apply a value added model to the data first.

Undoubtedly, the English system does not yet have the high drop out rates of the US system. However, the drop out rate has risen in recent years in the UK and the evidence here suggests that we should be alert to the fact that this will tend to widen

the socio-economic gap in degree completion, since poorer students drop out to a greater extent even after allowing for their prior achievement.

8. Reference

Altonji, J.G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics* **11** 1, pp. 48–83.

Philippe Belley & Lance Lochner, 2007. "The Changing Role of Family Income and Ability in Determining Educational Achievement," NBER Working Papers 13527, National Bureau of Economic Research, Inc.

Bennett, R. (2003). Determinants of undergraduate student drop out rates in a university business studies department. Journal of Further and Higher Education, 27(2), 123-141.

Blanden, J. and P. Gregg (2004), *Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain*, Oxford Review of Economic Policy, Vol. 20, pp. 245-263

Blanden, J. and S. Machin (2004), *Educational Inequality and the Expansion of UK Higher Education*, Scottish Journal of Political Economy, Special Issue on the Economics of Education, Vol. 51, pp. 230-49

Bound, J. and Turner, S. (2007) *Cohort Crowding: How Resources Affect Collegiate Attainment*, Journal of Public Economics 91, Issues 5-6, (June), pp. 877-899.

Callender, C. (2003), *Student Financial Support in Higher Education: Access and Exclusion*, in M. Tight (ed.) *Access and Exclusion: International Perspectives on Higher Education Research*, Elsevier Science, London

Carneiro, P. and J. Heckman (2002), *The evidence on credit constraints in post-secondary schooling*, Economic Journal, Vol. 112, pp. 705-734.

Carneiro, P. and J. Heckman (2003), *Human capital policy* in J. Heckman, A. Krueger and B. Friedman (eds.), *Inequality in America: What role for human capital policies?*, MIT Press, Cambridge, MA

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and A. Vignoles (2008). Understanding the determinants of participation in higher education and the

quality of institute attended: analysis using administrative data. Institute of Fiscal Studies, working paper.

CMPO (2006), *Family background and child development up to age 7 in the Avon Longitudinal Survey of Parnets and Children (ALSPAC)*, DfES Research Report No. RR808A

Cunha, F. and J. Heckman (2007), *The Technology of Skill Formation*, American Economic Review, Vol. 92, No. 2, pp.31-47

Cunha, F., J. Heckman, L. Lochner and D. Masterov (2006), *Interpreting the evidence on life cycle skill formation* in E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Chapter 12, pp.697-812, North-Holland, Amsterdam

Dearing, R. (1997). Higher education in the learning society: National Committee of Inquiry into Higher Education. London: HMSO.

Feinstein, L. (2003), *Inequality in the Early Cognitive Development of British Children in the 1970 Cohort*, Economica, Vol. 70, No. 277, pp. 73-98.

Ferrer-i-Carbonell, A. (2005). Income and well-Being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, 89, 997-1019.

Geronimus, A., Bound, J and Neidert, L. (1995) , 'On the validity of using Census geocode characteristicsto proxy individual socio-economic characteristics', NBER Technical Working Paper No. 189, National Bureau of Economic Research, Massachusetts, USA.

Hartog, J., Pfann, G. and Ridder, G. (1989). (Non-)Graduation and the earning function. An inquiry on self-selection. *European Economic Review* **33**, pp. 1371–1395.

Haveman, Robert & Wolfe, Barbara, 1993. "Children's Prospects and Children's Policy," *Journal of Economic Perspectives*, American Economic Association, vol. 7(4), pages 153-74, Fall.

Haveman, R. and Wilson, K. (2005) Economic Inequality in College Access, Matriculation and Graduation, mimeo. Paper presented to the Maxwell School Conference on Economic Inequality and Higher Education: Access, Persistence and Success, September 23-24.

HEFCE (1999) *Performance Indicators in Higher Education*, Guide 99/66, Bristol, HEFCE.

HEFCE (2005), *Young Participation in Higher Education*, HEFCE, Bristol (see: www.hefce.ac.uk/pubs/hefce/2005/05_03/05_03.pdf)

Johnes, G., and McNabb, R. (2004). Never give up on the good times: student attrition in the UK. *Oxford Bulletin of Economics and Statistics*, 66, 23-47.

Johnes, J. and Taylor, J. (1989) Undergraduate non-completion rates: differences between UK universities. *Higher Education*, 18, pp.209-225.

Machin, S. and A. Vignoles (2004), *Educational Inequality: The Widening Socio-Economic Gap*, Fiscal Studies, Vol. 25, pp. 107-28

Manski, C.F. (1989). Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review* 8, pp. 305–312.

Marcenaro-Gutierrez, O., Galindo-Rueda, F. and Vignoles, A. (2007) 'Who actually goes to University?', *Empirical Economics* (special issue "The Economics of Education and Training"), Vol. 32, No. 2, pp. 33-357.

McGivney, V. (1996). Staying or Leaving the Course: Non-Completion and Retention of Mature Students in Further and Higher Education. National Institute of Adult Continuing Education, Leicester, UK.

National Audit Office (2007) Staying the course: The retention of students in higher education, London: The Stationary Office.

National Audit Office (2008) Widening Participation in Higher Education, Report by the Comptroller and Auditor General, HC 725 Session 2008-2008, 25 June 2008, London: The Stationary Office.

Palmer, J. (2001) "Student Drop-out: a case study in new managerialist policy", *Journal of Further and Higher Education*, 25(3), 349-357.

Quinn, J. (2004) Understanding Working-Class 'Drop-out' from Higher Education through a Socio-cultural Lens: Cultural narratives and local contexts, International Studies in Sociology of Education, 14(1), 57-74.

Smith, J.P., & Naylor, R.A. (2001). Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *Journal of Royal Statistical Society: Series A*, 164, 389-405.

Turner, S. (2008) Common Across the Atlantic: The Under-representation of Low-Income Students in Higher Education in the U.S. and the U.K., mimeo produced for The Sutton Trust Conference on Social Mobility, 1 June 2008.

Wilson, D., S. Burgess and A. Briggs (2005), *The Dynamics of School Attainment of England's Ethnic Minorities*, CMPO Working Paper No. 05/130.

Yorke, M. (1998) Non-Completion of Undergraduate Study: Some Implications for Policy in Higher Education, *Journal of Higher Education Policy and Management*, v20 n2 p189-201.

Appendix Table 1: The socio-economic gradient in HE non-continuation by gender and university type

	Females	Males	Good university = 0	Good university = 1
Social class: Professional occupations	-0.0069	-0.0044	-0.0066	-0.0037
	[0.0022]**	[0.0028]	[0.0029]*	[0.0017]*
Social class: Associate professional and technical occupations	0.0019	0.0005	0.0020	0.0009
	[0.0027]	[0.0033]	[0.0032]	[0.0023]
Social class: Administrative and secretarial occupations	-0.0038	0.0050	0.0013	-0.0009
	[0.0031]	[0.0041]	[0.0038]	[0.0021]
Social class: Skilled trades occupations	0.0045	0.0032	0.0062	0.0006
	[0.0032]	[0.0039]	[0.0037]	[0.0023]
Social class: Personal service occupations	0.0054	0.0056	0.0102	-0.0009
	[0.0043]	[0.0070]	[0.0054]	[0.0050]
Social class: Sales and customer service	0.0073	0.0219	0.0213	0.0007
	[0.0055]	[0.0084]**	[0.0069]**	[0.0055]
Social class: Process, plant and machine	0.0059	0.0041	0.0041	0.0058
	[0.0046]	[0.0046]	[0.0047]	[0.0046]
Social class: Elementary occupation	-0.0079	0.0110	-0.0004	0.0013
	[0.0038]*	[0.0057]	[0.0053]	[0.0039]
Social class: Missing	0.0209	0.0230	0.0288	0.0130
	[0.0042]**	[0.0049]**	[0.0047]**	[0.0064]*
Multiple deprivation index	0.0028	0.0015	0.0035	0.0007
	[0.0010]**	[0.0010]	[0.0011]**	[0.0008]
Multiple deprivation index squared	-0.0009	0.0000	-0.0009	0.0000
	[0.0004]*	[0.0004]	[0.0005]	[0.0003]
Multiple deprivation index missing	-0.0064	0.0189	0.0113	-0.0787
	[0.0360]	[0.0409]	[0.0316]	[0.0189]**
OA education index	-0.0478	-0.0703	-0.0739	-0.0357
	[0.0217]*	[0.0217]**	[0.0239]**	[0.0215]
OA education index squared	0.0385	0.0655	0.0640	0.0321
	[0.0291]	[0.0293]*	[0.0320]*	[0.0259]
OA education index missing	-0.0319	-0.0492	-0.0570	0.6151
	[0.0242]	[0.0104]**	[0.0128]**	[0.0036]**
Free school meal	0.0036	-0.0010	0.0012	0.0022
	[0.0036]	[0.0034]	[0.0037]	[0.0035]
Observations	68976	52124	76077	45486
Log likelihood	-12867	-9561	-17728	-4774
Pseudo R-squared	0.1766	0.1891	0.1565	0.1566

Note: Same specification as the last column of Table 2. Standard errors in parentheses.

* indicates statistically significant at the 5% level. ** indicates statistically significant at the 1% level.

Appendix Table 2: The socio-economic gradient in HE non-continuation by ethnicity

	White	Black	Asian
Social class: Professional occupations	-0.0050	-0.0141	-0.0041
	[0.0021]*	[0.0168]	[0.0052]
Social class: Associate professional and technical occupations	0.0002	0.0104	0.0015
	[0.0024]	[0.0214]	[0.0078]
Social class: Administrative and secretarial occupations	-0.0001	0.0173	-0.0037
	[0.0028]	[0.0236]	[0.0071]
Social class: Skilled trades occupations	0.0020	0.0062	0.0100
	[0.0029]	[0.0331]	[0.0066]
Social class: Personal service occupations	0.0073	0.0195	-0.0133
	[0.0044]	[0.0283]	[0.0057]*
Social class: Sales and customer service	0.0155	-0.0248	0.0104
	[0.0059]**	[0.0132]	[0.0103]
Social class: Process, plant and machine	0.0042	0.0071	0.0039
	[0.0039]	[0.0330]	[0.0056]
Social class: Elementary occupation	-0.0004	-0.0287	-0.0053
	[0.0044]	[0.0107]**	[0.0055]
Social class: Missing	0.0235	0.0255	0.0122
	[0.0041]**	[0.0213]	[0.0055]*
Multiple deprivation index	0.0031	-0.0061	0.0020
	[0.0009]**	[0.0040]	[0.0013]
Multiple deprivation index squared	-0.0004	0.0012	-0.0003
	[0.0004]	[0.0016]	[0.0005]
Multiple deprivation index missing	0.0035	-0.0536	0.0018
	[0.0231]	[0.0131]**	[0.0517]
OA education index	-0.0377	-0.1163	-0.0106
	[0.0204]	[0.1132]	[0.0327]
OA education index squared	0.0080	0.1232	-0.0006
	[0.0277]	[0.1395]	[0.0439]
OA education index missing	-0.0246		
	[0.0192]		
Free school meal	0.0037	0.0077	-0.0007
	[0.0038]	[0.0110]	[0.0035]
Observations	87682	2321	11034
Log likelihood	-17453	-467	-1465
Pseudo R-squared	0.1719	0.2871	0.1792

Note: Same specification as the last column of Table 2. Standard errors in parentheses.
* indicates statistically significant at the 5% level. ** indicates statistically significant at the 1% level.