1 Introduction

As in the U.S. and many other countries, the underrepresentation of lower-income (or disadvantaged) students in colleges and universities has been an issue of great policy concern in the UK for many decades. Recent evidence for England indicates that the 20 percent most disadvantaged students are approximately six times less likely to go to a university compared to the 20 percent most advantaged (HEFCE 2005). Furthermore, the socioeconomic 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 U.S. (Cunha et al. 2006), much of the root cause of this inequality is located earlier in the educational system. Chowdry et al. (2008) have shown that in England if a disadvantaged student reaches a sufficient level of achievement in secondary school he has a similar chance of going on to university as a more advantaged student.

However, in both the UK and the U.S., the participation rate of lower-income students is not the only policy concern. To fully reap the rewards of a university education, students need to complete their education and attain degrees. Currently, around three-fourths of those who start their university education in the UK actually complete it—one-fourth drop out—this is a cause for concern. 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. One reason may be that disadvantaged students have less complete information about the real costs of studying for a degree. We know that in the UK lower-income students leave university with more debt and are more risk averse (Pennell and West 2005). This might mean that the students' socioeconomic backgrounds affect their likelihood of dropping out even if the family backgrounds do not affect university participation (Chowdry et al. 2008; Marcenaro-Gutierrez et al. 2007). Certainly, in the literature, dropping out or noncompletion has been seen as particularly problematic for students from disadvantaged backgrounds (Dearing 1997; McGivney 1996; HEFCE 1999; Quinn 2004).[1]

We examine whether disadvantaged entrants to college have a higher probability of dropping out given their level of prior achievement. Does being disadvantaged mean that you can get into university, but you are more likely to struggle and eventually become a dropout.[2] If this is true, then the focus of a wider participation policy should be on facilitating degree completion rather than just encouraging participation.

[1] The socioeconomic gap in dropouts from college in the U.S. system is discussed extensively in Turner (2008).

[2] In this paper, we use the terms college, university and higher education (HE) interchangeably.

With our data, we cannot specifically identify what causes a student to drop out of university. That is, we cannot differentiate between students who fail to meet the necessary academic standards and are required to drop out from students who simply choose to withdraw. Certainly, different causes of dropout have different policy solutions. Indeed, there is literature, discussed later, that acknowledges why some dropout may be rational if students drop out when they find the costs (mental or monetary) exceed the potential benefits (Manski 1989; Hartog et al. 1989; Altonji 1993).

Although we cannot identify the causes of dropout in our data, we can identify the importance of the students' prior academic achievements as compared to their family backgrounds. This is important because if poor academic preparation were largely to blame for dropout the policy solution would include improved school-based preparation for university as well as academic remedial help rather than financial assistance. This is particularly valuable information for the UK because the socioeconomic gap in HE participation has been found to be largely attributable to the poor academic preparation of 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 the underrepresentation of disadvantaged students in higher education must be at least partially school focused, thus the academic achievement of disadvantaged children needs to be raised. Therefore, the question this paper addresses is whether dropout from university is largely attributable to weak academic preparation or family-background-related factors.

To address this, we use administrative data on an entire cohort of young people in England who potentially could enter university in 2004–2005 at age 18.[3] These data are unique in that they include information on each student's personal characteristics (e.g., ethnicity, date of birth and indicators of socioeconomic background), and they provide a complete record of each child's educational achievement from age 11 onwards.[4] This is the first time that such longitudinal data have been available to study the important issue of college dropout in the UK.

The paper starts with a brief description of the English university system. We then 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.

---

[3] We have data on all students in English state-funded schools who turned age 18 in 2003–2004, that is, more than half a million students. We excluded students in private schools. To the extent that such students have lower university dropout rates, we may be understating the socioeconomic gap in university dropout.

[4] In the UK, students 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.

2 The English Higher Education 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 (UCAS). Students have to apply to specific degree courses; that is, they have to preselect a specific degree subject in a specific institution. Hence, in contrast to the U.S. 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 different institutions. 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 generally not 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 bachelor's degree and, unlike the U.S., transfer between institutions is not usual (fewer than 3 percent switch institutions). This is partly because there is no agreed upon system of college credits so transfer is difficult. Students' progressions from one year to the next are subject to a minimum academic level of achievement. This means that students may drop out because they fail to attain the required academic standard (the extent to which they have to pass examinations to progress varies by institution, but all have some form of assessment) or they anticipate failure. Students are allowed to repeat classes that they have failed. However, there are limits (that vary by institution) to the number of times a student can retake a class. Students that continuously fail are not allowed to continue.

As in the U.S., in England, there are concerns that financial barriers may prevent students from enrolling in HE in the first place and later cause them to drop out (Callender 2003). However, the English system differs significantly from the U.S. system in terms of the costs of study. The English system is 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 taxpayer. This has changed in the last decade, and institutions can now charge English students up to £3,000 per annum (although foreign students pay more). These fees are means tested with lower-income students being fully exempt from fees. Repayment is also now on an income-contingent basis after graduation to minimize the disincentive effect on students from disadvantaged

socioeconomic backgrounds. Furthermore, students also have access to loans to fund their living costs during their course of study.

One feature of the English university system that is particularly important is the fact that, unlike the U.S. and countries such as New Zealand and Sweden, England has historically had low levels of student dropout from higher education (Dearing 1997; NAO 2007). Recent data suggest that 91.6 percent of full-time students starting university in 2004–2005 continued into their second year and 78.1 percent are expected to complete their degree (NAO 2007).[5] However, as the sector has expanded and the rate of noncompletion has risen (Johnes and McNabb 2004), policy attention has shifted to this issue, and noncompletion rates are now part of a range of indicators used by the government to measure university performance. Indeed, university league tables that rank universities on a number of metrics including the dropout rate are reported in UK newspapers. These league tables generally do not take into account student characteristics and, therefore, may provide a misleading impression of the true institutional quality in terms of retaining students, particularly for universities with a large number of disadvantaged and lower-achieving students. Ideally, value-added models, which control for students' prior achievements, are needed to assess whether institutions have particularly high or low dropout 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 operationalized using administrative data.

3 Previous Literature

There is a large and growing literature on the role of family background, that is, income and socioeconomic status in determining educational outcomes, particularly university achievement (Blanden and Gregg 2004; Carneiro and Heckman 2002, 2003; Gayle et al. 2002; Meghir and Palme 2005; Haveman and Wolfe 1993). Such studies have found family background to be an important determinant of educational achievement and have suggested that the socioeconomic 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 U.S.). In fact, Cunha et al. (2006) conclude that family background, specifically credit constraints, plays only a limited role in determining HE participation, conditional on achievement in secondary school, although some recent studies dispute this (Belley and Lochner 2007).

[5] The dropout rate in the data is somewhat lower than this because we focus specifically on young, full-time entrants. Comparable HESA data indicate that the UK noncontinuation rate from the first year of study to the second for young, full-time entrants was 7.2 percent in 2004–2005. In the data for England, the figure is very similar (just over 6 percent).

Even if educational inequality emerges early, as it appears to in the U.S., the raw socioeconomic gap widens substantially if one measures bachelor degree completion as opposed to enrollment (Turner 2008). This raises the question as to whether the conditional dropout rate is higher for lower socioeconomic students even when taking into account their prior educational achievement. The literature on the relationship between socioeconomic background and dropout from university is limited, although the U.S. evidence finds differential dropout by family income (see Haveman and Wilson 2005; and related issues in Bound et al. 2007).

In England, Johnes and McNabb (2004) analyzed students entering and leaving the "old" (pre-1992) universities and distinguished between voluntary and involuntary (failure) dropout.[6] Jones and McNabb found that students from a lower socioeconomic background were more likely to drop out voluntarily. Smith and Naylor (2001) used the same data to examine completion and noncompletion. Using a binomial regression analysis of the probability that an individual withdraws from university for whatever reason, Smith and Naylor found the risk of dropping out to be very high for students from lower-social-class backgrounds and those 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 a degree course in the Business Studies department of a "new" university in Greater London. Using data he collected, Bennett estimated a confirmatory factor analysis of the probability of noncompletion among business students. He found self-declared financial hardship to be one of the strongest predictors of an individual's decision to withdraw. Other important factors included low self-esteem and inferior academic performance at university. While these studies were able to control to some extent for a student's entry qualifications, they did not have rich data on the students' prior achievements and only considered a subset of UK universities.

This paper estimates models of student dropout or noncompletion for all universities in England. We are concerned with student dropout for two main reasons. First, there may be economic costs associated with noncompletion (Yorke 1998). Second, a student may experience a sense of failure after dropping out of university, which might affect his future productivity and earnings. Thus, we must be cautious in interpreting our results.

Although the UK does not have the long-standing U.S. tradition of part-time and intermittent higher education, increasingly, UK students do enroll part-time, and their progression may not be linear. This is particularly true for mature students (McGivney 1996). Labeling (temporary) withdrawal as an academic

[6] 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.

failure or wastage seems to be inappropriate because although students may withdraw from their studies it does not mean that they have not received any benefit (Johnes and Taylor 1989). This is not merely a semantic debate. Universities in England now have an incentive to encourage student completion in the "normal" time because noncompletion whatever the cause is financially penalized.[7] If disadvantaged students are more likely to drop out than their more advantaged counterparts for a given level of prior achievement, this could lead to tension between the widening participation agenda that encourages greater participation by disadvantaged students and the desire by universities not to incur penalties because of high levels of student withdrawal (Palmer 2001).

Some economists have also made the argument that dropout from HE is efficient—weaker students who would not benefit from completing their degrees appropriately drop out. Manski (1989), for example, argues that lower dropout rates would not necessarily make society better off. He suggests that the decision to enroll is a decision to initiate an experiment, a possible outcome of which is dropout (see also Hartog et al. 1989; Altonji 1993). Thus, enrollment in HE incurs a risk for all students, namely the risk that they may have to drop out for whatever reason. Disadvantaged students may face higher levels of risk. For example, they may be more likely to fail to reach the level of educational achievement required, or they may make decisions about the choice of institution and subject of study based on inferior information. This higher level of risk may partially explain lower participation rates by disadvantaged students. Even if disadvantaged students face the same risk of dropout as their advantaged peers, if they are more risk averse (Callender 2003), then this too would at least partially explain their lower enrollment 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).[8] The school administrative data (NPD/PLASC) contain each student's record of his primary and secondary schooling, including information on educational attainment from age 11 to 18, personal characteristics, such as date of birth, ethnicity, home postcode, entitlement to free school meals and whether English is an additional language in their home.[9] The university

[7] Public funding received by each university in England is potentially affected by their performance. Noncompletion is one measure of university performance used by the authorities.

[8] These data are maintained by the Department for Children, Schools and Families.

[9] This can be thought of as a proxy for very low family income. Students 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–

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,[10] we have longitudinal data on a cohort of students from age 11 through to potential HE participation at age 18 in 2004–2005 and continued participation at age 19 in 2005–2006. For 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 the student continued from one year to the next, that is, continued at the same university from 2004–2005 to 2005–2006. We recognize that we are only able to consider dropouts from the first year of study, and, if students from different socioeconomic 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 socioeconomic gap in dropout in comparison to what we might expect to see in terms of degree completion. Around 6 percent of the students failed to progress from one year to the next, indicating that they dropped out (voluntarily or involuntarily) or decided to transfer to another institution the following year. As our cohort only potentially entered HE the previous year, we are essentially measuring dropout after the first year of three years of study.[11] With our data, we are not able to differentiate those who drop out of higher education altogether from those who switch institutions. Data from the Higher Education Statistics Agency suggest that 2.9 percent of students from low-income neighborhoods switch institutions (across all years of enrollment) as compared to 2.7 percent of students from higher-income neighborhoods.[12] Since switching occurs at a similar rate across different types of students, 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 in English, mathematics and science given to all children in English state schools. The tests are externally validated; that is, individuals not associated with the schools score them. We include test scores in three subject areas, English, mathematics and science, using a nonlinear specification. The test data are supplemented by the results from public examinations taken by most students at

---

2008 prices).

[10] 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.

[11] The vast majority of UK bachelor's degrees are three years of study.

[12] Higher Education Statistics Agency: http://www.hesa.ac.uk/dox/performanceIndicators/0607/t3b_0607.xls .

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, which gives the total number of points accumulated from the student's eight highest GCSE grades.[13] At 18, we use the total (uncapped) point score. We divide the population into five evenly sized quintile groups, ranked according to their scores at GCSE and A-level or equivalent, to capture attainment at these levels.[14] These data contain the best possible information on students' prior achievements, which enables us to identify the distinct roles that academic preparation and socioeconomic background play in dropping out of university.

Based on the student's university, we also linked in an institution-level indicator of the university's research quality from the 2001 Research Assessment Exercise (RAE).[15] We then combine this indicator of the quality of each institution's research with an indicator of whether 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, recognizing the inherent problems in defining what constitutes quality in higher education from a student's perspective.

The data have a number of limitations. First, the indicators for students' family backgrounds 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 not available. Further, if we rely on data about parental socioeconomic background for young, full-time entrants to a first degree, we have missing data for around one in five students. This is a well recognized problem in the UK. For instance, the UK National Audit Office has stressed the poor quality of data available to identify the socioeconomic 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 a

[13] We use a capped total point score to avoid conflating the quantity of GCSEs taken and the grades received because 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), although we may not believe these are equivalent in terms of achievement.

[14] 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.

[15] 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.

student was eligible for free school meals (FSM) in secondary school. This indicator is limited by not being contemporaneous with HE participation and also by measuring the very bottom of the income distribution. Around 5 percent of students entering HE was eligible for free school meals in secondary school. We also have each student's postcode (similar to a zip code in the U.S.) in secondary school when they were applying to enter university, and we can use this geographical marker to link information on the characteristics of the student's neighborhood.[16] Specifically, we can link measures of socioeconomic deprivation (namely, an Index of Multiple Deprivation (IMD) score, derived from Census data on the characteristics of individuals living in their neighborhood,[17] and an Income Deprivation Affecting Children Index (IDACI) score, again constructed based on Census data on individuals living in their neighborhood).[18] While these administrative measures of family deprivation are not ideal (family income would be preferable), we use individual and administrative measures of socioeconomic background and test the robustness of our findings by using these alternative indicators. In Table 1, we include individual measures of parental socioeconomic 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 socioeconomic circumstances of those students for whom we have missing individual level data.

Another limitation of the data is that we only consider young HE participants, that is, those participating at age 18. The dropout behavior of older HE participants may differ, so our results cannot be generalized to include older students. Finally, we only have data on state school students. A significant minority of students in England attend private schools prior to entering HE (just under 7 percent at age 16 in our data). If these more advantaged students were included in our sample and if they have very low dropout rates, then our estimates of the socioeconomic gap in HE dropout may well be the lower bounds.

In Table 1 below, we show full descriptive statistics for our sample, including a comparison with students who do not enroll in university (column 1) and those who entered HE one year later at age 19 (final column). Nonparticipants are more likely to be white males. However, what is most striking is that only 6

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

[17] The IMD score is available at the 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.

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

percent of the sample of nonparticipants achieved the top quintile of GCSE achievement (at age 16) while around half of those enrolling in higher education reached this higher level of achievement. At age 18, again the nonparticipants had much lower levels of achievement, primarily because most of them left the educational system by that age and did not take A-levels or high school graduation equivalent qualifications (thereby getting a zero score). Nonparticipants are also more likely to be disadvantaged when measured by their free school meal status and their geographical-based gauge of deprivation.

Moving from left to right across Table 1, we consider the sample who enrolled in university at age 18. From this sample, dropouts 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.

While we cannot include those who enter HE at 19 in 2005–2006 in our analysis (because we do not have data on these students for 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, the late entrants appear more likely to be non-white, lower achievers, 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 noncontinuation as follows:

$$\begin{aligned} &\Pr(Y = 1 \mid X = x) = \Phi(x'\beta), \\ &Y_{ij} = X_{ij}'\beta + \varepsilon_{ij}, \end{aligned} \tag{1}$$

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 socioeconomic 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 ages 11, 14, 16 (GCSE) and 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 socioeconomic 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 disadvantaged students are more likely to drop out of university, allowing for their prior educational achievement. While we are confident that we can fully control for each student's prior achievement, which previous research has been unable to do, we recognize that there may be unobserved factors, which determine whether a student drops out of university and that can be correlated with their socioeconomic status. For example, disadvantaged students may tend to enroll in certain types of universities. There may also be unobserved features of a student's university that determine their decision to withdraw. Although we control for degree subject and institutional quality in a relatively comprehensive way, there are many other features of universities that we are unable to accommodate, such as pastoral care, advice and guidance given, and so on.

Table 1: Descriptive Statistics of Selected Variables					
	Nonparticipants	***All HE participants***	***Continued to year 2***	***Dropped out year 1***	***Age 19 entrants***
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

Table 1: Descriptive Statistics of Selected Variables					
	Nonparticipants	***All HE participants***	***Continued to year 2***	***Dropped out year 1***	***Age 19 entrants***
	(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

Table 1: Descriptive Statistics of Selected Variables					
	Nonparticipants	***All HE participants***	***Continued to year 2***	***Dropped out year 1***	***Age 19 entrants***
	(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 Advanced level or equiv point score	24.335	288.365	295.875	169.512	193.977
	(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)	

Table 1: Descriptive Statistics of Selected Variables					
	Nonparticipants	***All HE participants***	***Continued to year 2***	***Dropped out year 1***	***Age 19 entrants***
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)	

Table 1: Descriptive Statistics of Selected Variables

	Nonparticipants	***All HE participants***	***Continued to year 2***	***Dropped out year 1***	***Age 19 entrants***
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.

To partially address this, we estimate models that allow for mean differences in dropout across different institutions. We do this by including a dummy variable for each university, which allows for any differences in the noncontinuation 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 equations:

$$\Pr(Y = 1 \mid X = x) = \Phi(x'\beta), (1)$$

$$Y_{ij} = X_{ij}'\beta + \lambda_{j+}\varepsilon_{ij}, (2)$$

We estimated an HE participation equation for everyone in our sample and obtained a predicted likelihood of HE participation. We included this predicted probability in the dropout equation. The variable measuring the individual's predicted probability of HE participation was significantly negatively correlated with whether 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 dropout 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 dropout decisions.[19]

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

6 Results

Table 2 below estimates the model described above, where the dependent variable takes a value of one if the student continues beyond his first year of study in the same university and zero otherwise. We then sequentially add various explanatory variables as controls. Column 1 controls for the student's socioeconomic background and individual characteristics only. The probability of dropping out varies by socioeconomic background. In broad terms, students with higher socioeconomic backgrounds who live in less materially deprived and more educated neighborhoods have lower dropout rates. It is noticeable that the geographically based measures remain significant even with the inclusion of individual level measures of socioeconomic background. The fifth of the sample

[19] Results available on request.

who do not have an individual level indicator of socioeconomic background (i.e., who have a missing dummy variable for socioeconomic status) are significantly (by nine percentage points) more likely to drop out of university than students from a managerial/senior official background. For those students with missing data, the only indicators of socioeconomic background that we have are the geographical measures.

To illustrate the magnitude of socioeconomic differences in dropout, a student 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 student whose parents are in sales or customer service is around three percentage points more likely to drop out than a student from a managerial/senior official background. These socioeconomic differences are large in the context of low overall dropout (6 percent in our sample).

In terms of other individual characteristics, there appears to be an ethnicity dimension to the problem of dropout. Almost all 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 white students (Wilson et al. 2005). Moreover, many ethnic minority students are also more likely to attend college in the UK even without allowing for prior educational achievement (Chowdry et al. 2008). In our data, Black African, Pakistani and Bangladeshi students are approximately three percentage-points less likely to withdraw than white students. Chinese students are around four percentage-points less likely to drop out compared to students of white ethnic backgrounds. Other individual characteristics, such as gender, are not significant.

Column 3 in Table 2 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 the individual achieved an A-level or high school graduation equivalent qualification (level 3).[20] While 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 dropout. What is striking, however, is the strong relationship between prior achievement at age 16 (GCSE) and 18 (A-level) with dropout. When we include these prior achievement variables, the coefficients of the socioeconomic variables are reduced, although most of the socioeconomic 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 sales or customer service are now two percentage points more likely to drop out than students from a managerial/senior official

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

background. This result implies that not all the socioeconomic gradient in dropout is attributable to inferior prior academic achievement of disadvantaged pupils.

In column 4 of Table 2, we control for degree subject and institution status. The institution status and degree subject variables are insignificant, and there are only marginal reductions in the coefficients on the socioeconomic variables.[21] As discussed earlier, we also want to control for the unobserved characteristics of universities that may affect student dropout. We do this by including institution dummy indicator variables in Column 5. The inclusion of these dummy variables does not change the coefficients of the socioeconomic variables in a major way; that is, there remains a socioeconomic gradient in university dropout even after controlling for prior attainment and institution means. This suggests that the higher withdrawal rate of disadvantaged students is not necessarily attributable to the types of degree subjects they are studying and the institution type in which they enroll. In any case, since the students' choices of degrees and institutions is itself related to their socioeconomic backgrounds, this specification removes some of the socioeconomic effect we are trying to measure.

Although we control for prior achievement in the models, we also explored interactions between prior achievement and noncontinuation. We therefore reestimated the model separately for students in the top three quintiles of the age-11 test score distribution, as shown in Table 3. The general finding from Table 2 holds; that is, there remain socioeconomic differences in the tendency for students to drop out even after controlling fully for prior achievement. However, the individual level socioeconomic background variables are largely 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 five 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. Students with a parent in a professional occupation are less likely to drop out than students with a parent in a managerial occupation. In fact, the magnitude of the coefficient is, if anything, greater for students in the third quintile of age-11 achievement. However, the coefficients are insignificant and the standard errors are large. This is partially driven by the smaller numbers of students with professional parents who achieve in the third quintile of age-11

[21] Note that in models without measures of prior achievement, there are large statistically significant differences in the likelihood of dropout 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 dropout across degree subject, controlling only for socioeconomic 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.

achievement. In contrast, students in the third quintile of age-11 achievement who have a parent in sales or customer service are three percentage points more likely to drop out than students in that quintile of achievement with managerial parents. This result is significant, but it does not hold for students in the top quintiles of age-11 achievement.

Further interactions by gender and by university status and ethnicity were explored and are reported in the appendix. It is difficult to discern clear patterns because some (but not all) of the socioeconomic background variables become insignificant when estimated separately by subgroup. While the coefficients on some variables do vary across the different groups of students and institutions, the results suggest that the relationship between socioeconomic background and student dropout is quite similar by gender and institution type (with a marginally smaller socioeconomic gradient for students at higher status institutions and a higher socioeconomic gradient for white students, but not significantly so). For black students in contrast, there appears to be very little socioeconomic gradient in dropout once you control fully for prior achievement. One motivation for this paper was to explore whether universities have different dropout rates once we control for the prior achievement of their students. This is important. The policy importance of university league tables in the UK and the notion of dropout rates are metrics 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 dropout 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, 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 specific institutions (institution F), the inclusion of student characteristics and prior achievement make a significant difference to the magnitude of the coefficient. In other words, for some institutions, it may be that raw dropout rates can be misleading and 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, for example, ethnicity, because we do not want to argue that dropout rates vary inherently by this particular individual characteristic. That said, it is useful to potential students and universities to consider value-added adjustments to dropout rates.

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	-0.0130	-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	0.0018	0.0015	0.0007	0.0002
	[0.0030]	[0.0025]	[0.0024]	[0.0023]
Social class: Skilled trades occupations	0.0119	0.0056	0.0044	0.0039
	[0.0031]**	[0.0025]*	[0.0024]	[0.0023]
Social class: Personal service occupations	0.0204	0.0079	0.0054	0.0057
	[0.0061]**	[0.0045]	[0.0039]	[0.0038]
Social class: Sales and customer service	0.0306	0.0166	0.0139	0.0127
	[0.0069]**	[0.0050]**	[0.0047]**	[0.0046]**
Social class: Process, plant and machine	0.0113	0.0059	0.0051	0.0049
	[0.0040]**	[0.0035]	[0.0034]	[0.0034]
Social class: Elementary occupation	0.0123	0.0027	0.0008	0.0002
	[0.0052]*	[0.0040]	[0.0037]	[0.0035]
Social class: Missing	0.0929	0.0403	0.0244	0.0222
	[0.0175]**	[0.0084]**	[0.0057]**	[0.0038]**
Multiple deprivation index	0.0041	0.0011	0.0024	0.0022
	[0.0017]*	[0.0013]	[0.0009]*	[0.0007]**

Table 2: The socio-economic gradient in HE non-continuation				
Multiple deprivation index squared	-0.0011	-0.0005	-0.0006	-0.0005
	[0.0004]**	[0.0004]	[0.0003]*	[0.0003]
Multiple deprivation index missing	0.0190	-0.0032	0.0018	0.0019
	[0.0328]	[0.0201]	[0.0222]	[0.0202]
OA education index	-0.1112	-0.0621	-0.0704	-0.0595
	[0.0237]**	[0.0196]**	[0.0183]**	[0.0165]**
OA education index squared	0.1066	0.0763	0.0735	0.0516
	[0.0311]**	[0.0261]**	[0.0248]**	[0.0216]*
OA education index missing	-0.0727	-0.0489	-0.0494	-0.0416
	[0.0105]**	[0.0106]**	[0.0092]**	[0.0097]**
Free school meal	0.0011	-0.0003	0.0012	0.0018
	[0.0029]	[0.0027]	[0.0026]	[0.0024]
Blacks	-0.0178	-0.0248	-0.0236	-0.0242
	[0.0052]**	[0.0029]**	[0.0025]**	[0.0020]**
South Asians	-0.0323	-0.0295	-0.0263	-0.0243
	[0.0032]**	[0.0023]**	[0.0019]**	[0.0013]**
Chinese	-0.0429	-0.0348	-0.0332	-0.0322
	[0.0041]**	[0.0028]**	[0.0026]**	[0.0021]**
Mixed ethnicity	-0.0269	-0.0197	-0.0195	-0.0185
	[0.0051]**	[0.0045]**	[0.0043]**	[0.0040]**
Other ethnicity	-0.0169	-0.0205	-0.0193	-0.0198
	[0.0051]**	[0.0032]**	[0.0029]**	[0.0025]**
Unknown ethnicity	-0.0012	-0.0013	-0.0011	-0.0008
	[0.0034]	[0.0027]	[0.0026]	[0.0026]

Table 2: The socio-economic gradient in HE non-continuation				
Male	0.0027	-0.0015	-0.0021	-0.0012
	[0.0024]	[0.0017]	[0.0015]	[0.0014]
Date of birth	0.0000	0.0000	0.0000	0.0000
	[0.0000]	[0.0000]*	[0.0000]*	[0.0000]**
Disabled	0.0084	-0.0040	-0.0059	-0.0080
	[0.0054]	[0.0037]	[0.0031]	[0.0021]**
Level 3 at 18		-0.0322	-0.0229	-0.0195
		[0.0081]**	[0.0043]**	[0.0038]**
Level 3 at 18: Missing		-0.0148	-0.0051	0.0030
		[0.0141]	[0.0176]	[0.0197]
Subject: Missing			0.1908	0.1558
			[0.0385]**	[0.0339]**
Subject: Allied to medicine			-0.0090	-0.0082
			[0.0084]	[0.0082]
Subject: Biological science			-0.0017	-0.0036
			[0.0097]	[0.0090]
Subject: Veterinary science			0.0119	-0.0030
			[0.0173]	[0.0128]
Subject: Agriculture			0.0423	0.0096
			[0.0243]	[0.0139]
Subject: Physics			-0.0035	-0.0030
			[0.0084]	[0.0082]
Subject: Mathematics			-0.0025	-0.0002
			[0.0082]	[0.0083]

Table 2: The socio-economic gradient in HE non-continuation		
Subject: Computer science	0.0060	0.0027
	[0.0111]	[0.0102]
Subject: Engineer & Technology		0.0059
	[0.0114]	[0.0110]
Subject: Architecture	-0.0020	-0.0043
	[0.0101]	[0.0091]
Subject: Social studies	0.0038	0.0013
	[0.0101]	[0.0093]
Subject: Law	-0.0013	-0.0039
	[0.0094]	[0.0086]
Subject: Business	0.0094	0.0038
	[0.0110]	[0.0098]
Subject: Mass communication		0.0159
	[0.0168]	[0.0103]
Subject: Languages	0.0039	0.0018
	[0.0099]	[0.0094]
Subject: History	-0.0025	-0.0032
	[0.0107]	[0.0103]
Subject: Arts	0.0291	0.0155
	[0.0166]	[0.0121]
Subject: Education	0.0004	-0.0051
	[0.0106]	[0.0089]
Subject: Combined	0.0101	0.0107
	[0.0128]	[0.0126]

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

Good university			0.0036	
			[0.0065]	
Institutional indicator variables	No	No	No	Yes
Key stage test scores	No	Yes	Yes	Yes
P-values				
Test of joint significance – KS3 English		0.106	0.025	0.017
Test of joint significance – KS3 Maths		0.001	0.001	0.001
Test of joint significance – KS3 Science		0.561	0.261	0.290
Test of joint significance – KS4		0.000	0.000	0.000
Test of joint significance – KS5		0.000	0.000	0.000
Observations	121827	121827	121827	121591
Log likelihood	-25700.72	-23836.47	-23080.39	-22586.77
Pseudo R-squared	0.0641	0.132	0.1595	0.1771

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 percent level. ** indicates statistically significant at the 1 percent 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 percent level.

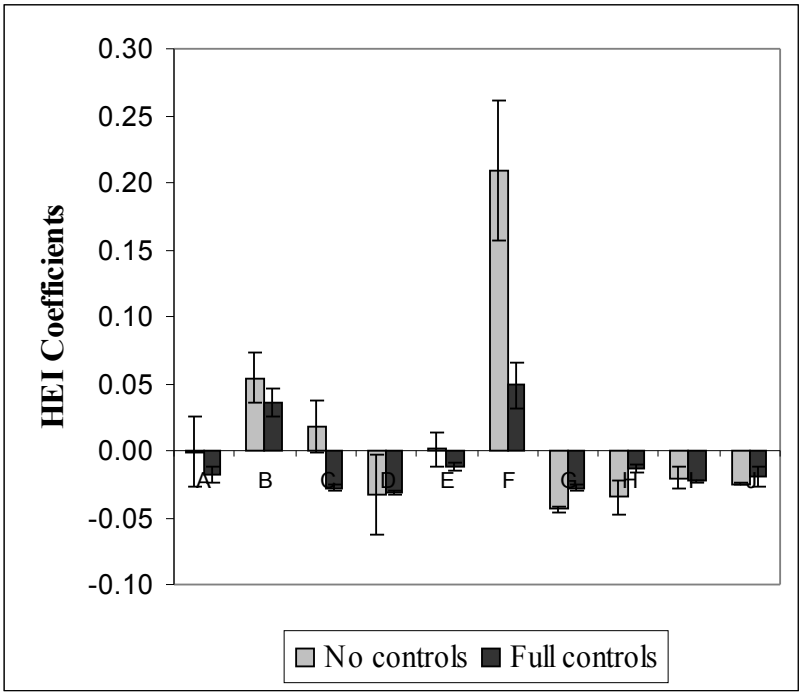
Table 3: Socioeconomic Gradient in HE Noncontinuation by Key Stage 2 Achievement

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

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	4.3e+07
Total number of institution coefficients per equation	119
N of institutions that we can reject the equality test at the 5 percent level, i.e., t>=1.96	67
percent 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 four standard-error bars (two above and two below) to indicate 95 percent C.I.

7 Conclusions

Our results suggest that there exists a significant gap in the dropout rate between advantaged and disadvantaged students in the UK. Much of this gap disappears when we allow for students' prior achievements, suggesting that some of the apparent difference in first-year dropout rates between advantaged and less-advantaged students is actually attributable to differences in their academic preparation for HE or their ability as measured by earlier standards of educational achievement. Given the relatively low aggregate rate of first-year dropout (6

percent) from English universities, once we fully control for prior achievement, a student 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 student whose parents are in sales or customer service 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 dropout rate in the UK, these differences in dropout rates are arguably sizeable.

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

We also found that raw indicators of the dropout 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' dropout rates can indeed change markedly in individual cases if the prior achievement of students is fully taken into account. In policy terms, this suggests that if we were to use dropout rates as measures of institutional performance, we must be careful to apply a value-added model to the data first.

Obviously, the English higher educational system does not yet have the high dropout rates of the U.S. system. However, the dropout rate has risen in recent years in the UK, and the evidence suggests that we should be alert to the fact that this tends to widen the socioeconomic gap in degree completion because disadvantaged students drop out to a greater extent even after allowing for their prior achievement.

Appendix

Table A1: Socioeconomic Gradient in HE Noncontinuation by Gender & 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 percent level. ** indicates statistically significant at the 1 percent level.

Table A2: The Socioeconomic Gradient in HE Noncontinuation 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	0.0002	0.0104	0.0015
	[0.0024]	[0.0214]	[0.0078]
Social class: Administrative and secretarial	-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 percent level. ** indicates statistically significant at the 1 percent level.

References

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

Belley, Philippe, and Lance Lochner. 2007. The Changing Role of Family Income and Ability in Determining Educational Achievement. NBER Working Paper 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* 20:245–263.

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* 51:230–49.

Bound, J., M. Lovenheim, and S. Turner. 2007 Understanding the Decrease in College Completion Rates and the Increased Time to the Baccalaureate Degree. PSC Research Report No. 07-626. November 2007.

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

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

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

Chowdry, H., C. Crawford, L. Dearden, A. Goodman, and A. Vignoles. 2008. Understanding the determinants of participation in higher education and the quality of institute attended: analysis using administrative data. Working paper Institute of Fiscal Studies.

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* 92 (2): 31–47

Cunha, F., J. Heckman, L. Lochner, and D. Masterov. 2006. Interpreting the evidence on life cycle skill formation. In *Handbook of the Economics of Education,* eds. E. Hanushek and F. Welch, 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* 70 (277): 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.

Gayle, V., D. Berridge, and R. Davies. (2002) Young people's entry into higher education: quantifying influential factors. *Oxford Review of Education* 28: 5–20.

Geronimus, A., J. Bound, and L. Neidert. 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., G. Pfann, and G. Ridder. 1989. (Non-)Graduation and the earning function. An inquiry on self-selection. *European Economic Review* **33:**1371–1395.

Haveman, Robert, and Barbara Wolfe. 1993. Children's Prospects and Children's Policy. *Journal of Economic Perspectives*, American Economic Association, vol. 7 (Fall): 153–74.

Haveman, R., and K. Wilson. 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.

———. 2005. Young Participation in Higher Education. Bristol, HEFCE, www.hefce.ac.uk/pubs/hefce/2005/05_03/05_03.pdf.

Johnes, G., and R. McNabb. 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 J. Taylor. 1989. Undergraduate non-completion rates: differences between UK universities. *Higher Education* 18:209–225.

Machin, S., and A. Vignoles 2004. Educational Inequality: The Widening Socio-Economic Gap. *Fiscal Studies* 25:107–128.

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

Marcenaro-Gutierrez, O., F. Galindo-Rueda, and A. Vignoles. 2007. Who actually goes to University? *Empirical Economics* (special issue "The Economics of Education and Training") 32 (2): 333–357.

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.

Meghir, C., and M. Palme. (2005) Educational reform, ability, and family background. *American Economic Review* 95: 414–424.

National Audit Office. 2007. *Staying the course: The retention of students in higher education.* 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.

Pennell, H. and A. West. (2005) The impact of increased fees on participation in higher education in England. *Higher Education Quarterly.* 59: 127–137.

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., and R.A. Naylor. 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* 20 (2): 189–201.