Research Report RW92

Ethnicity and Degree Attainment

Stijn Broecke and Tom Nicholls
Department for Education and Skills

department for
education and skills
creating opportunity, releasing potential, achieving excellence

Ethnicity and Degree Attainment

Stijn Broecke and Tom Nicholls
Department for Education and Skills

Contents

Ethnicity and Degree Attainment[1]

Summary

Although the participation of students from minority ethnic communities in Higher Education (HE) is higher than for students from White communities, the attainment of those who complete a first degree programme (as measured by class of degree) is markedly lower than that of their White peers.

Previous studies have shown that this difference in attainment cannot be explained by age, prior attainment, or subject of study – but, as far as we know, there has been no large scale analysis of recent cohorts which has systematically tried to control for a full range of factors in analysing the attainment of students from minority ethnic communities in HE.

In this study, we seek to address this gap in the literature. We select 65,000 qualifiers from the 2004/05 Higher Education Statistics Agency (HESA) data to include English-domiciled undergraduate qualifiers who started their course of study in 2002/03. This enables us to have information about these students' prior attainment (in the form of tariff scores) as well as their socio-economic background (proxied through the Index of Multiple Deprivation) – whilst still including a large number of students from minority ethnic communities.

We use a Partial Proportional Odds Model (a special form of Ordered Logistic Regression) to predict the likelihood of obtaining a certain degree class based upon a number of characteristics. In essence, this approach allows us to compare the odds different groups of students face in obtaining a certain class of degree.

In our model, we are able to include the following variables to predict HE attainment: prior attainment, subject of study, age, gender, disability, deprivation, type of HE institution attended, type of level 3 qualifications, mode of study, term-time accommodation, and ethnicity.

Variables we were not able to include were: type of prior institution attended, term-time work, parental income and other parental attributes, and English as an additional language.

Our results show that, even after controlling for the majority of factors which we would expect to have an impact on attainment, being from a minority ethnic community (except the "Other Black", "Mixed" and "Other" groups) is still statistically significant in explaining final attainment, although the gap has been significantly reduced.

Care must be taken in interpreting these results and in drawing conclusions. In particular the following points need to be taken into account:

- o Although we were able to control for a large number of characteristics, we certainly could not include the whole range of factors which could impact on the performance of students from minority ethnic communities in HE.

- o Although we included tariff scores and type of qualification (vocational/academic) in our analysis, together these are still not a full control for prior attainment.

- o It should be remembered that we only included qualifiers in our analysis, i.e.: we only included those students who actually completed their course of study and excluded all those who did not complete the programme[2].

- o Apart from the identification of Russell Group institutions, our model did not allow the difficulty of getting a good degree to vary by institution or subject within an institution. However, previous work suggests that any such differences are small.

Our prime interest was in the explanatory power of the ethnicity variables and so we spent relatively little time exploring the plethora of other results that this study provided us with. However, some of the most interesting other findings were:

- o That men performed significantly less well than women - except when it came to getting firsts, when there was no statistically significant difference between the two sexes.

- o That students entering with academic rather than vocational level 3 qualifications tend to do better.

- o That students at Russell Group institutions were more likely to perform well, except when it came to getting firsts (when the Russell Group effect became insignificant).

- o That students living at home were more likely to get firsts.

- o And that the likelihood of getting a good degree increased with age, albeit with diminishing returns.

---

[2] Some work by HEFCE suggests that students from minority ethnic communities have similar completion rates to students from White communities after entry qualifications, subject and age are taken into account. However, studies have shown that men do have lower completion rates than women, after taking into account a range of factors.

I - Introduction

1. The purpose of this paper was to improve our understanding of the complex issue of performance in Higher Education (HE) in relation to students from minority ethnic communities. In particular, the analysis tried to shed light on whether being from a minority ethnic community plays a significant role in explaining the attainment of those who complete a first degree programme (as measured by class of degree), once other factors determining attainment are controlled for. Our main objective was not an exact quantification of the impact of ethnicity (and other factors) on attainment, but rather to understand the direction of their effect (if significant), as well as their relative importance.

2. The report contains: a brief discussion of previous analysis of degree attainment and ethnicity; our choice of methodology to model degree attainment; a description of the data used; results from the econometric analysis; and a final section which summarises the findings and draws some conclusions.

II - Previous analysis[3]

3. The HE participation rates of students from minority ethnic communities tend to be higher than those of their White peers. In "Why the Difference?" Connor et al (2004) reported Higher Education Initial Participation Rates (HEIPRs) for English domiciled first-time entrants (full- and part-time) to HE courses (in universities and colleges), by broad ethnic/gender group for the academic year 2001/02:

Table 1: HEIPR by Ethnic/Gender group

Ethnic Group	Male	Female	All
White	34	41	38
All Minority Ethnic	55	58	56
- Asian or Asian British	62	59	60
- Black or Black British	55	66	61
- Chinese or Other Ethnic	47	50	49
- Mixed Ethnic	35	44	40
All (With Known Ethnicity)	37	43	40

Source: Connor et al (2004)

4. However, despite high HE participation rates, a number of studies have found that students from minority ethnic communities perform less well than their White peers: Connor et al. (2003 p.74 -77), using Higher Education Statistics Agency (HESA) data for the 1998/99 academic year, found that minority ethnic students appear less successful in attaining upper second or first class degrees (so-called "good" degrees) relative to White groups and, in "Why the Difference?", Connor et al. (2004 p.75-79) used 2001/02 HESA data for England to show that students from all minority ethnic communities (apart from the small

---

[3] This section is not intended to provide an exhaustive literature review – it merely presents a background to the study by illustrating some of the main findings encountered in the literature concerning ethnicity and HE attainment that sparked our interest in carrying out the analysis presented in this paper.

"mixed" and "other" ethnic groups) are less than half as likely as students from 'White' communities to obtain a good degree:

Table 2: Degree Attainment by Ethnic Group

Ethnic Group	1st	2.1	2.2	3rd or lower	Total (n)
White	10.7	48.9	31.0	9.4	157,741
All Minority Ethnic	6.0	36.2	40.1	17.8	28,292
Black Caribbean	2.9	32.0	48.6	16.5	2,016
Black African	3.3	27.5	45.6	23.5	3,275
Black Other	4.7	34.2	43.2	17.9	1,092
Indian	6.6	38.5	38.7	16.2	8,837
Pakistani	4.5	34.5	41.9	19.0	3,813
Bangladeshi	3.7	35.6	42.2	18.5	1,183
Chinese	8.9	37.6	36.1	17.5	2,127
Asian Other	7.2	36.5	34.7	21.5	2,349
Mixed Ethnic	9.4	47.0	35.5	8.1	330
Other	8.5	41.5	35.9	14.1	3,270
All (Known Ethnicity)	10.0	47.0	32.4	10.7	186,033

Source: Connor et al (2004)

5. A limitation of this kind of information is that it does not control for other factors which may impact on attainment, so that the 'ethnicity' effect identified may simply be a result of other differences between White students and minority ethnic students which impact on attainment. From the literature (e.g. Purcell et al. (2005), Connor et al (1996), Connor et al. (2004)) we know that a number of factors other than ethnicity impact on HE attainment: "The Class of '99" (Purcell et al. (2005), for instance, finds that prior attainment is an important factor in predicting degree outcome, as well as gender, school characteristics, university and subject studied. Other findings from the same study indicate that those who are reported to be non-disabled, describe themselves as 'ambitious' and have a father with a degree, are all also more likely to attain a good class of degree.

6. In fact, Connor et al. (2004 p.75-79) believe that most of the differences in HE attainment between ethnic groups can be ascribed to these other factors, and they recommend that a more technical analysis be carried out to control for some of these in order to estimate the "pure" effect of ethnicity and the response to it upon degree attainment (if it exists). Connor et al. (2003) mention some preliminary research by the Higher Education Funding Council for England (HEFCE) and HESA that has tried to do this – but they still find some 'unexplained difference' in attainment between different ethnic groups. Purcell et al. (2005) similarly find indications that those who are classed as White are more likely to obtain a 'good' degree, even after controlling for some other characteristics.

7. One of the reasons these studies still find an "ethnicity component" to attainment may be that they only control for some of the influences on degree performance, and that econometric analysis including a more complete set of explanatory variables would be required – something we seek to address in this study.

8. Section V of this paper discusses the factors we would expect to have an impact on HE attainment, and which ones of these we were able to include in our econometric model as explanatory variables. First, however, we describe the econometric approach we took to model degree attainment.

III - Methodology

9. In choosing the appropriate methodology for our analysis, we were forced to go beyond simple Ordinary Least Squares (OLS) analysis because of the non-continuous (or discrete) nature of our dependent variable, degree class: an individual's degree outcome is limited to the set of values "1st", "2.1", "2.2" or "3rd/pass" and cannot take any of the values in between.

10. Discrete dependent variables normally call for **logistic regression** which enables us to estimate the likelihood of an outcome occurring given different values of the independent variables. Even here, however, different approaches are possible. HEFCE (2003), for instance, run a series of binary regressions with different outcomes (getting a 2.2 or better; getting a 2.1 or better; getting a first). However, it may be more parsimonious to group such a series of binary regressions into a single regression model with **multiple outcomes**.

11. In addition, because the order of the outcomes in our regression matters (a 1st is better than a 2.1 which is better than a 2.2 etc…), we opted for **ordered logistic regression** as our preferred methodology. Ordered logistic regression suffers from a major drawback, however, which is that it assumes that the (very restrictive and often violated) **proportional odds (or parallel lines) assumption** holds. In order to understand what this means, we first need to explain how the output from ordered logistic regression can be interpreted.

12. Although logistic regression in Stata outputs the results in **ordered logits** (or ordered log odds), which are difficult to interpret, we can add a simple option to the Stata command so that the results are expressed in proportional **odds ratios** instead – which are much more intuitive. An odds ratio gives us the ratio, given a one-unit increase in the explanatory variable, of the odds[4] of being in a higher rather than a lower category. For instance, let's assume that we are trying to predict degree class based on gender. Degree outcome is defined by the set {3rd, 2.2, 2.1, 1st} and gender by the set {0, 1}, where "0" is male and "1" female. The Stata output from running this regression would look something like:

---

[4] Odds are just an alternative way of expressing the likelihood of an event such as catching the flu. Probability is the expected number of flu patients divided by the total number of patients. Odds would be the expected number of flu patients divided by the expected number of non-flu patients.

class	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
3						
gender	1.5	.0732202	14.42	0.000	1.66073	1.948053
2.2						
gender	1.3	.026269	20.48	0.000	1.398811	1.501805
2.1						
gender	1.0	.0282472	0.39	0.697	.957073	1.067855

13. Each horizontal line in the output corresponds to a different degree outcome: the first corresponds to a 3rd, the second to a 2.2, and the third to a 2.1. The way we interpret the "odds ratio" column is as follows: the odds ratio in the first line (i.e. the one corresponding to a 3rd) compares the odds of obtaining anything better than a 3rd (i.e. the odds of obtaining a 2.2, a 2.1 or a 1st rather than a 3rd) when gender increases by 1 (i.e. for females as compared to males). In this particular example, the odds ratio is 1.5: the odds of getting a 2.2 or better rather than a 3rd are 150 for females to 100 for males (re: the odds measure ratio of the probability of an event occurring to the probability of its not occurring). Similarly, the second line compares the odds for females as compared to males of getting a 2.1 or a 1st instead of a 2.2 or 3rd. In this example, the odds of females are 130 to 100 for males. Finally, there is no difference between females and males when it comes down to the likelihood of obtaining a 1st as opposed to anything below that (last line in the output).

14. We are now in a position to explain what the proportional odds assumption entails: ordered logistic regression is only valid if the odds ratios are identical for all possible outcomes. In the example above, this entails that we would only be entitled to use ordered logistic regression if the odds ratios were identical for all outcomes; i.e. if females' likelihood of getting a better degree was the same irrespective of what degree outcomes we were comparing. Although this assumption may hold in some cases, it is easy to see how restrictive it is, and how easily it may be violated.

15. We have tested all of our variables for whether they violate the proportional odds assumption using a Brant Test. The results indicate that some variables do violate this assumption, whereas others don't. Usually, this implies that we should abandon our model and use an alternative, like the multinomial regression model, instead. The problem with such models is that they include far more parameters than the proportional odds model, which complicates interpretation. This is because they free all variables from the proportional odds assumption, even though the assumption may only be violated by a few.

16. As a solution to this problem, we decided to use a **partial proportional odds model** using the program Gologit2 written by Richard Williams, where the parallel lines assumption is relaxed only for those variables where it is not justified. This method will first test all our independent variables, and then estimate the model holding some variables constant across dependent outcomes and allowing others to vary,

providing us with a partial proportional odds model. The variables that are found to hold for the proportional odds assumption will have constant coefficients over the range of outcomes, whereas the variables that violate this assumption will report different coefficients for each of the three reported outcomes[5].

IV - Data and Coverage

17. We used the 2004/05 HESA qualifiers dataset for our analysis. We reduced this dataset by applying a number of filters to include only:

Qualifiers with level 3 qualifications as their highest qualification upon entry to HE. This enabled us to model prior attainment on a consistent scale using the UCAS tariff system and reduced our dataset from 341,412 to 186,067 records. Deleting all cases where the tariff score was missing resulted in a further reduction from 186,067 to 93,250.

Qualifiers who started their course in 2002/03. Although the tariff score had been collected for some candidates prior to 2002/03, this has only been consistently done since 2002/03. We therefore deleted all records of students who entered HE prior to 2002/03, reducing our dataset from 93,250 to 78,110 individuals. Note that, since the HESA data we worked with was the 2004/05 dataset, this meant that our final dataset only included full-time students who completed their course in 3 years.[6]

Qualifiers who attained a degree classification. This does not just filter out those who did not complete their degree, but also some studying subjects where a degree classification is not given, such as in most Medicine and Veterinary Science degrees. This filter resulted in a further reduction from 78,110 to 74,614 cases.

English domiciled students. Because only for these qualifiers did we have Index of Multiple Deprivation (IMD) data[7]. This resulted in a reduction from 74,614 to 67,734 records.

Those where the age was realistic (older than 13 and younger than 99 years) **and where IMD data wasn't missing**. This resulted

---

[5] Variables that are held constant over the three outcomes in our model are: all of the ethnic groups; disability; the tariff score; the "non- A level" and "A level mix" dummies; the log of age; the Index of Multiple Deprivation (IMD) rank; and some of the subjects (Combined, Languages, Veterinary Science, and Education). These were identified using the autofit function of the gologit2 command at the 1% significance level.

[6] Within the data there is a very small minority (less than 0.05%) that is registered as part-time. This must be an anomaly, and we left these students in the dataset as full-time students.

[7] The IMD measures deprivation for every Super Output Area and local authority area in England. It combines indicators across seven domains into a single deprivation score and rank. The domains are: income deprivation; employment deprivation; health deprivation and disability; education, skills and training deprivation; barriers to housing and services; living environment deprivation; and crime.

in a further reduction from 67,734 to 67,546.

Those with known level 3 qualifications. Unless we knew a qualifier's full set of level 3 qualifications, we removed them from our analysis. This was to ensure that the qualifier only had A levels, a mix of A levels and vocational qualifications, or solely vocational qualifications upon entry. This final filter reduced our dataset from 67,546 to 66,649 records – the size of our final dataset.

18. One advantage of applying the filters in the way we did (i.e. include only full-time first degree students who entered with level 3 qualifications) is that we ended up with a set of relatively homogenous individuals. In a way, then, the filters we used can be seen as a first "control" for other factors which impact on degree attainment: although our analysis will be carried out on a small subset of all students in the HESA record, we end up comparing "like for like". In addition, although the total size of the dataset was reduced from 340,000 to 65,000 cases, this still provided us with a sizeable dataset for the purposes of our analysis, including a large number of students from minority ethnic communities:

Table 3: Ethnic Minority Composition of All versus Selected Qualifiers

	All Qualifiers		Selected Qualifiers	
Ethnic Group	**Frequency[8]**	**Proportion (%)**	**Frequency[9]**	**Proportion (%)**
White UK & Irish	256,285	75.1	55,165	82.8
Other White background	6,595	1.9	845	1.3
Black or Black British - Caribbean	3,945	1.2	520	0.8
Black or Black British - African	7,365	2.2	815	1.2
Other Black background	1,275	0.4	110	0.2
Asian or Asian British - Indian	11,230	3.3	3,255	4.9
Asian or Asian British - Pakistani	5,640	1.7	1,375	2.1
Asian or Asian British - Bangladeshi	1,760	0.5	475	0.7
Chinese	2,860	0.8	845	1.3
Other Asian background	3,950	1.2	715	1.1
Total Mixed	4,325	1.3	1,335	2.0
Other Ethnic background	2,480	0.7	410	0.6
Total unknown/Refused	33,700	9.9	785	1.2
Total	341,410[10]	100	66,650[11]	100

19. As table 3 indicates, the total proportion of the White UK & Irish group was larger in our final dataset than in the original dataset (83% v. 75%) but this can be attributed mainly to the reduction in the "unknown/refused" group. Another notable difference between the two datasets is the reduction in the relative size of the Black or Black British ethnic groups, whereas the Asian or Asian British groups are now overrepresented. A possible explanation for this could be the way the

---

[8] Rounded to the nearest 5.
[9] Rounded to the nearest 5.
[10] May not add up due to rounding.
[11] May not add up due to rounding.

reduced dataset has been constructed. We used only those who have level 3 qualifications as their highest qualification upon entry to HE. However, as table 4, below illustrates students from Black groups are more likely to enter HE with other qualifications, so that they would have been taken out from our dataset. Asian groups, on the other hand, are more likely than the average to enter with level 3 qualifications:

### Table 4: Ethnicity and Type of Entry Qualification

Ethnic Group	Level 3 Qualifications	Other Qualifications
White UK & Irish	59%	41%
Other White	39%	61%
Black or Black British – Caribbean	42%	58%
Black or Black British - African	38%	62%
Other Black	37%	63%
Asian or Asian British – Indian	76%	24%
Asian or Asian British – Pakistani	73%	27%
Asian or Asian British – Bangladeshi	72%	28%
Chinese	70%	30%
Other Asian	51%	49%
Total Mixed	68%	32%
Other Ethnic	54%	46%
Total unknown/refused	56%	44%
Total	54%	46%

20. As predicted, the filters we applied to the original dataset reduced the attainment gap between students from minority ethnic communities and students from White communities (see figures 1 and 2 below):

Figure 1: Attainment Gaps, All Qualifiers
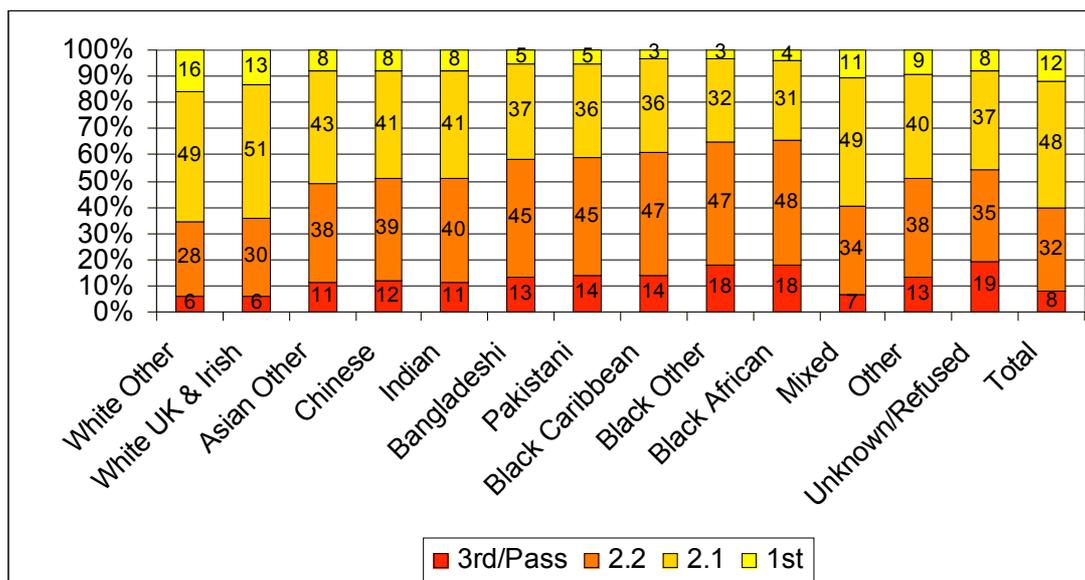
Figure 2: Attainment Gaps, Selected Qualifiers
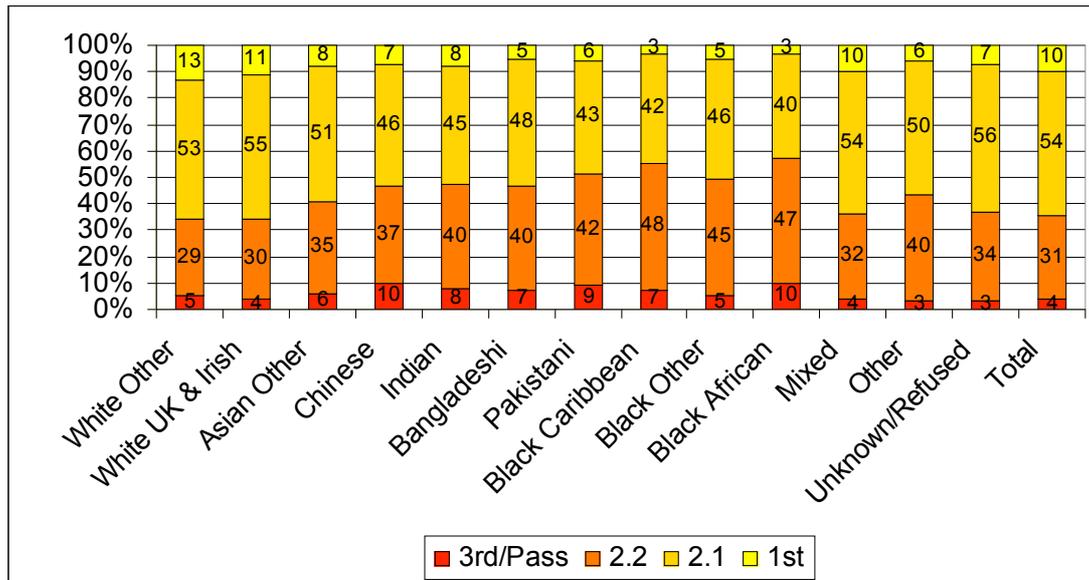
V - The model

21. In this section we discuss: the factors which we would expect to have an impact on attainment; which ones we were able to include in our model as explanatory variables; how we coded them; and which variables we were, unfortunately, unable to control for.

22. The factors we would expect to affect attainment in HE include:

> **Prior attainment** - where higher prior attainment should provide an indication of the ability of the qualifier. This could be in the form of **tariff scores** or **type of level 3 qualification** (vocational qualifications may equip individuals with different skill sets which could affect their performance in HE - this could be particularly relevant for students from minority ethnic communities if they have a greater propensity to enter HE with vocational qualifications). Looking at the position in aggregate, minority ethnic students have lower entry qualifications on average and fewer take an A level route compared with White UK & Irish students.

> The type of **prior institution** may also help to explain how someone performs in HE. Certain institutions may or may not equip people with the necessary skills to help attain higher degree classifications. Conversely, some institutions may cause some to 'over-achieve' at level 3 which could weaken the relationship between prior attainment and degree outcome.

> **Parental attributes** such as **parental income** – those students whose parents or guardians provide them with more financial support may not need to undertake term-time work (see below) and will have more resources which may help their attainment in their HE course (purchase of books, better study facilities, etc…). **Parental education** is another factor we would expect to impact

positively on achievement in HE.

Gender - there are reported gender gaps at all levels of the educational system, with women generally performing better than men.

Age - if students from minority ethnic communities are more likely to follow non-traditional routes into HE, then they are likely to be older than their White peers. Age could in itself have an effect on attainment, but the direction of effect is hard to predict.

Disability - those with disabilities may also have different attainment to those who are reported to have none. This could be due to access and participation issues.

English as an Additional Language - those whose mother tongue is not English may struggle more in HE due to language barriers.

Subject of study - would affect degree outcome as grading varies from subject to subject. There is also evidence of a concentration of minority ethnic students in particular HE subjects, e.g.: medicine, law, engineering, mathematics, business and computer science.

Type of HE institution (HEI) attended - minority ethnic student participation is generally weighted in favour of the post-92 universities. It may be that certain groups of HEIs achieve different degrees of value added, resulting in different proportions of "good" degrees attained.

Term-time working - evidence suggests that there is a link between term-time working and degree attainment (*Class of '99*) and that students working 'full time' were more likely to be from minority ethnic communities (amongst others) (HEFCE, 2005).

Other interesting variables which could have been investigated include the **mode of study** (full- or part-time – though this dataset only includes full-time students) and **term-time accommodation**. Both of these may have potential findings for other policy areas.

23. In an ideal world, we are able to control for all variables that we would expect to impact on attainment. In practice, however, we were only able to include a number of the above-mentioned variables in our model. These are:

Ethnicity - which we included at sub-group level: White (UK & Irish), Other White (non-UK & non-Irish), Black African, Black Caribbean, Other Black, Asian Bangladeshi, Asian Indian, Asian Pakistani, Asian Chinese, Other Asian, Mixed, Other and Not Known (including refused). This variable is a simple flag (i.e. we include dummies for each ethnic sub-group), where White (UK &

Irish) is the reference group.

Subject area - we included dummy variables for the following subjects: Medicine & Dentistry, Subjects Allied to Medicine, Biological Sciences, Veterinary Science, Agriculture & Related Subjects, Physical Sciences, Mathematical Sciences, Computer Science, Engineering & Technology, Architecture, Building & Planning, Law, Business & Administrative Studies, Mass Communications, Languages, Historical and Philosophical Studies, Creative Arts & Design, Education, Combined and Social Studies. This latter subject was chosen as the reference category since its ethnic composition was similar to that of the dataset including all qualifiers.

Prior Attainment - only first degree qualifiers with level 3 qualifications on entry, and those who entered in 2002/03 have been included within our dataset. This allows us to use the UCAS tariff[12] (a common currency for all level three qualifications) to control for prior attainment across all types of level 3 qualifications. This variable is used as a continuous variable, using the entire range of points. We divide the tariff score by 10 so the coefficient reflects larger changes in prior achievement.

Type of level 3 qualification - we have used those with A levels as our reference group, and flagged students who came in with a mix of A levels and vocational qualifications, and those with just vocational qualifications.

Index of Multiple Deprivation - the rank of the Index of Multiple Deprivation (IMD) is used to model relative deprivation. We have used IMD as we do not at present have any information on parental income on the HESA dataset. IMD is a composite area based measure of deprivation (see footnote 7 for more detail) which is derived using postal code information on the area of the qualifier's registered permanent home. A ranked variable of IMD was created (1-10), where 1 is the highest level of deprivation and 10 is the lowest level. One weakness of using this measure to proxy for parental income is that the student may not necessarily be deprived him or herself, but that their permanent home happens to be in an area of relative deprivation. In addition, this measure should not be interpreted as equivalent to a measure of socio-economic classification (SEC); it will encapsulate some of the same things as a SEC measure but ultimately it is a measure of relative deprivation[13].

---

[12] There are some concerns about the quality of the UCAS tariff score as a measure of prior attainment, and it is possible that if a better measure had been available, the unexplained 'ethnicity' effect may have been further reduced.

[13] SEC was tested within our model but IMD had better coverage; there was more of our subset which had IMD data relative to SEC. IMD was also found to have greater explanatory power.

Gender - gender is modelled with males as the reference group and females as a dummy.

Disability - disability is captured by flagging those who are registered on the HESA record as having a disability (even though they are not necessarily in receipt of a Disabled Student Allowance). No distinction is made between different types of disability due to small cell sizes.

Age - age is modelled as a continuous variable, but in a logarithmic form. This was found to be the best fit for the age variable and allows for a diminishing impact of age on attainment as the individual gets older.

Institution attended - the type of HE institution has also been modelled. There were a number of options ('New', 'Old', Other, Private and Russell Group), which are not all mutually exclusive. Russell Group HEIs were found to have the strongest explanatory power in terms of degree attainment, and so we chose to include a dummy for them, with non-Russell group HEIs as the reference group. The degree of correlation between Russell Group and entry requirement is considered not too high to prevent modelling[14].

Term-time accommodation - where the qualifier is reported to be living at his or her parents' or guardians' home, a dummy is created. This is modelled using information from the year of qualifying (information about term-time accommodation is collected each academic year, but we assume that the individual's term-time accommodation in the final year is representative of where s/he has been living throughout his/her degree).

24. As can be seen from the above list of variables, we were able to control for a substantial number of factors in our analysis of the impact of ethnicity on degree attainment. Unfortunately, there were a number of variables that we were not able to include in our model, including: **type of prior institution**, **term-time working**, **parental income** and **other parental attributes**, and **English as an Additional Language**. The significance of this will be discussed in the concluding section of this paper.

25. The final model we used was:

$$Degree_C = \alpha + Gender_D + Tariff + \text{Level 3 Type}_D + Disabled_D + Ethnicity_D +$$
$$\text{IMD rank} + Subject_D + \text{Russell Grp}_D + \text{At Home}_D + \log(age) + \varepsilon$$

Where: $_C$ = Third/Pass, 2:2, 2:1 or First
$_D$ = Dummy variable

---

[14] Pair wise correlation value of 0.47 (see correlation matrix of variables in Annex).

VI - Results[15]

26.   All model results are reported in the Annex. Due to the complexity of interpreting these results, we report the main findings here.

27.   Despite controlling for other factors which impact on attainment, we find that ethnicity is still statistically significant in explaining attainment in HE: all students from minority ethnic communities (except Other White, Other Black, Mixed and Other)[16] are found to be less likely to achieve a better degree relative to White UK & Irish students – and this result holds at all levels of attainment.

28.   In the two figures below, we: (i) repeat the original probabilities of attaining certain degree outcomes for our dataset of selected qualifiers (this is identical to Figure 2, except we now only include the ethnic groups that were still statistically significant in explaining degree outcome in our controlled analysis); and (ii) present the probabilities of attaining certain degree outcomes obtained by our model (i.e. after controlling for all the factors discussed in section V).

*Figure 3: Degree Classification by Ethnicity – Original Probabilities*

---
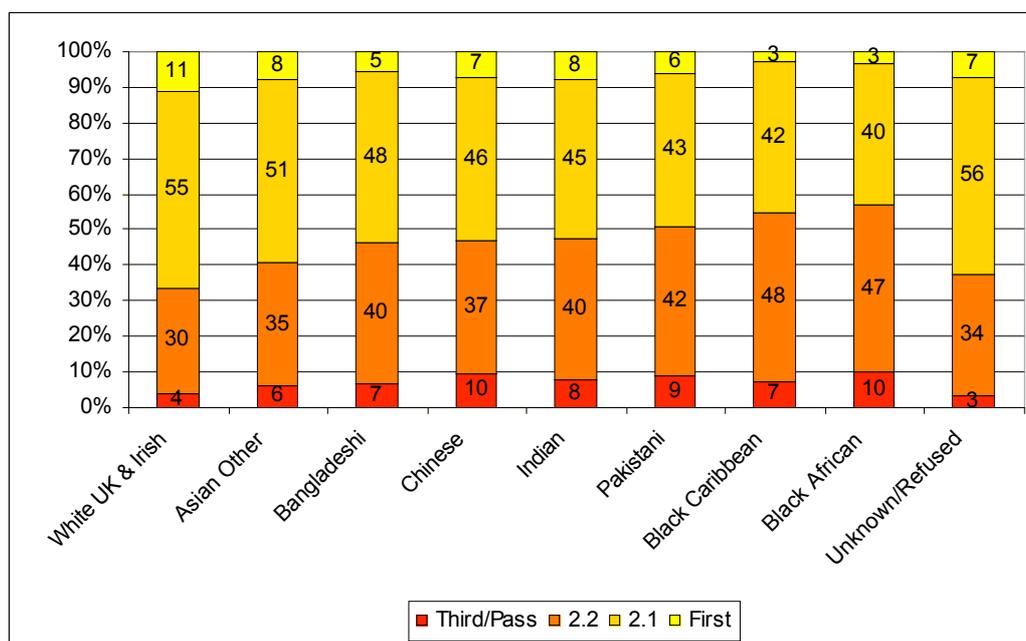
[15] All results reported within this section are significant at the 95% confidence interval unless stated otherwise.

[16] Bangladeshi qualifiers are found to perform less well; at a lower level of statistical confidence than the other ethnic groups modelled. However the group is still significant at the 95% confidence interval.

Figure 4: Degree Classification by Ethnicity – "Controlled"
Probabilities[17]
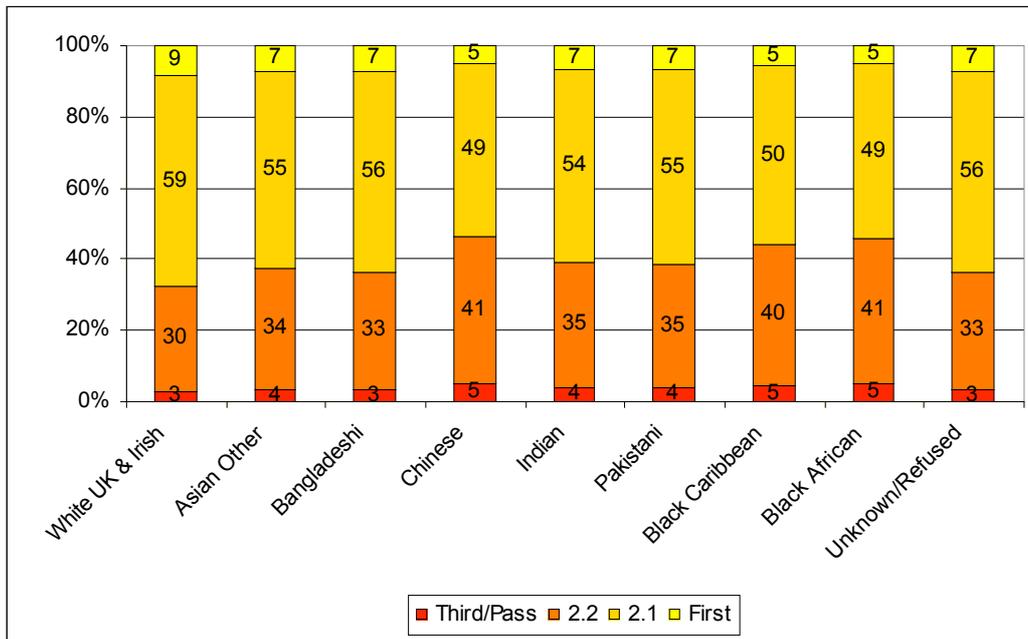
29. Comparing the controlled and uncontrolled analyses for each ethnic group, we find that the variables we included in our model help to explain some of the differences in attainment, but not the entire gap.

30. Although the primary aim of this analysis was to look at the effect of ethnicity on attainment, it is worth mentioning some of the other results briefly as well:

> **Prior attainment** at level 3 is found to have a very significant effect on degree outcome. The chart below shows the relationship between total UCAS tariff score and the probability of attaining different degree classifications. Unsurprisingly, the probability of obtaining a "good" degree increases significantly with the total number of tariff points obtained, even after controlling for other characteristics.

---

[17] For each ethnicity, these probabilities have been calculated by setting the value of the other ethnicity variables to zero, and the value of all other variables to their mean.

Figure 5: Degree Classification as a Function of Prior Attainment (Controlled Analysis)
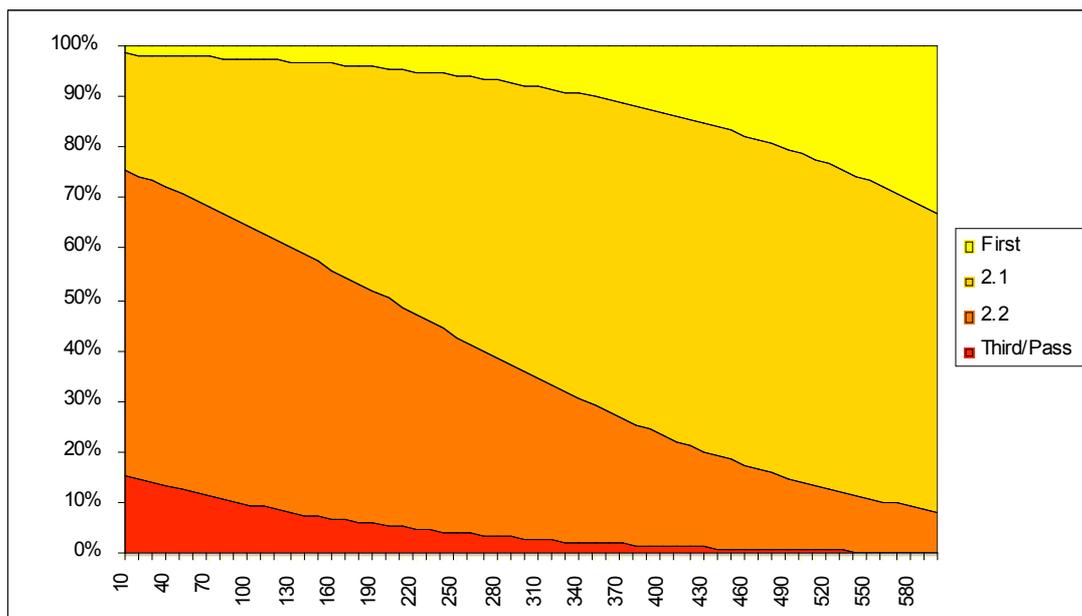
Note: Grade A at A level = 120pts

Type of level 3 qualification is found to be significant in explaining degree attainment. Qualifiers with a mix of A levels and vocational qualifications and those with no A levels are found to attain less well across the degree outcomes relative to those who only have A levels.

Gender is also found to be significant in explaining degree classification. Females are more likely to obtain a higher degree classification than males, except when it comes to attaining a first, where we found no statistically significant difference between the two sexes. The gender effect is strongest at the lowest outcomes.

Disability is found to reduce the probability of getting a better degree outcome.

As the **IMD deprivation** rank increases (i.e. as the individual's relative deprivation decreases), the probability of obtaining a better degree increases.

Subject of study also raised a number of significant results. Grossly simplifying and generalising, and relative to the Social Sciences, there is a HIGHER probability of obtaining a good degree in the subjects of Mass Communications, Languages and Historical and Philosophical studies, and a LOWER probability in the science subjects (except when it comes to attaining a first compared to a lower class of degree). In four subjects students were consistently less likely to attain a higher degree class than in the social sciences, and these were: medicine, veterinary sciences, law, and education.

Type of HEI was modelled by looking only at whether the qualifier was at a Russell group institution or not. This variable is found to increase the probability of attaining a higher degree, except when it comes to getting firsts, when the coefficient on the variable becomes insignificant.

Students **living at home** were more likely to attain a first. At all other degree outcomes, the variable was insignificant.

Age is found to increase the probability of a higher degree classification over the entire range of degree classifications, but the magnitude of this effect is reduced as age increases.

VII - Conclusion

31. The objective of this analysis was to find out whether ethnicity, after controlling for other factors, still had predictive power in explaining degree attainment. The main result of this report is that, after controlling for gender, prior attainment (tariff score and type of level 3 qualifications), disability, deprivation, subject of study, type of HEI, term-time accommodation, and age, there is still an unexplained difference between students from minority ethnic communities and students from White (UK and Irish) communities when we look at a subset of qualifiers who entered with level 3 qualifications. The attainment gap remains the largest for Black Caribbean, Black African and Chinese students.

32. These results potentially have quite serious implications. A number of studies have found that attaining a 'good' degree carries a premium in the labour market, and that this premium has been increasing over time, as the HE system has expanded. (Battu et al, 1999; Naylor et al. 2003; Bratti et al. 2005). As a result, there is a considerable cost attached to this attainment gap identified in relation to minority ethnic students.

33. Our findings do not automatically imply, however, that there is some form of ethnic bias within the HE system, and there are a number of other reasons which could explain the results we obtained. For one, there are other variables which we would have liked to include in our model, but for which no data was available. These include term-time working, parental income and education, English as an Additional Language, and prior institution attended. In addition, the controls we were able to include are not as sophisticated as they could be (e.g.: information on prior attainment). It seems reasonable to assume that, had we been able to include these variables (or improve the quality of those included), the attainment gap between White (UK and Irish) and minority ethnic students would have been further reduced. However, it is difficult to judge the extent of that reduction, or whether or not it would have been eliminated entirely.

34. Secondly, our analysis was performed on qualifiers only. If, for some reason, ethnic minority drop-outs had higher potential than White UK &

Irish drop-outs (in the sense that they would have attained better if they had stayed on), then this would further explain part of the attainment gap.

35.   It should be noted, though students from minority ethnic communities are less likely to perform well than their White counterparts, men perform significantly less well than women.

Stijn Broecke (Stijn.Broecke@dfes.gsi.gov.uk)
Tom Nicholls (Tom.Nicholls@dfes.gsi.gov.uk)

References

Battu, H., C. Belfield and P. Sloane (1999): "Overeducation among graduates: a cohort view", Education Economics, 7, 21-38.

Bhattacharyya, G., L. Ison and M. Blair (2003): "Minority Ethnic Attainment and Participation in Education and Training: The Evidence", DfES Research Topic Paper RTP01-03.
http://www.dfes.gov.uk/rsgateway/DB/RRP/u013881/index.shtml

Bratti, M., R. Naylor, J. Smith (2005): "Variations in the Wage Returns to a First Degree: Evidence from the British Cohort Study 1970", IZA Discussion Paper No. 1631, Institute for the Study of Labor, Bonn.

Connor H., I. La Valle, N.D. Tackey and S. Perryman (1996): "Ethnic Minority Graduates: Differences by Degrees", IES Report 309.
http://www.employment-studies.co.uk/pubs/report.php?id=309

Connor, H., C. Tyers, S. Davis and N.D. Tackey (2003): "Minority Ethnic Students in Higher Education: Interim Report", DfES Research Report RR448.
http://www.dfes.gov.uk/rsgateway/DB/RRP/u014352/index.shtml

Connor, Tyers, T. Modood and J. Hillage (2004) "Why the Difference? A Closer Look at Higher Education Minority Ethnic Students and Graduates", DfES Research Report RR552.
http://www.dfes.gov.uk/rsgateway/DB/RRP/u013556/index.shtml

HEFCE (2003): "Schooling effects on higher education achievement: Further Analysis- Entry at 19", HEFCE Issues Paper February 2005/09.
http://www.hefce.ac.uk/pubs/hefce/2005/05_09/

HEFCE (2002): "Survey of higher education students' attitudes to debt and term-time working and their impact on attainment".
http://www.hefce.ac.uk/Pubs/rdreports/2005/rd15_05/

Naylor, R., J. Smith and A. McKnight (2003): "Returns to Educational Performance: Evidence from UK graduates' first destination labour market outcomes", mimeo, University of Warwick, Coventry.

Purcell, K., P. Elias, R. Davies and N. Wilton (2005): "The Class of '99: A study of the early labour market experience of recent graduates", DfES Research Report RR691 RTP01-03.
http://www.hecsu.ac.uk/cms/ShowPage/Home_page/Research_reports/Class_of__99/p!eklFiai

Williams, Richard (2006), *Generalized Ordered Logit, Partial Proportional Odds Models for Ordinal Dependent Variables The Stata Journal* 6 (1):58-82. A pre-publication version is available at
http://www.nd.edu/~rwilliams/gologit2/gologit2.pdf

Annex: Model estimates

Generalized Ordered Logit Estimates

Log likelihood = -65221.29

Number of obs = 66649
LR chi2(73) = 12117.79
Prob > chi2 = 0.0000
Pseudo R2 = 0.0850

newclass2	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
gender	1.83581	.0755679	14.76	0.000	1.693517	1.990059
totalt~fby10	1.061289	.0008818	71.59	0.000	1.059562	1.063019
disabled	.9087724	.0276915	-3.14	0.002	.856087	.9647001
otherwhite~h	.9406713	.0659336	-0.87	0.383	.8199273	1.079196
blackcareth	.6038396	.0525064	-5.80	0.000	.5092207	.7160398
blackafreth	.5662486	.0396528	-8.12	0.000	.4936282	.6495527
otherblack~h	.794723	.1471667	-1.24	0.215	.5528285	1.14246
indianeth	.7532382	.0276392	-7.72	0.000	.7009685	.8094057
pakistaneth	.7638234	.0424348	-4.85	0.000	.6850208	.8516911
banglaeth	.8324394	.0756654	-2.02	0.044	.6965973	.9947717
chineth	.5609069	.0384146	-8.44	0.000	.4904501	.6414852
otherasian~h	.7940324	.0590554	-3.10	0.002	.6863267	.9186405
mixedeth	.9416231	.0519983	-1.09	0.276	.8450299	1.049258
othereth	.8577698	.0827088	-1.59	0.112	.7100603	1.036206
notknowneth	.8477365	.0601601	-2.33	0.020	.7376577	.974242
rank	1.017524	.0031734	5.57	0.000	1.011324	1.023763
medden	.1333084	.030269	-8.87	0.000	.0854247	.2080326
alliedmed	.7531187	.0788393	-2.71	0.007	.613418	.9246352
biosci	.8057016	.0603812	-2.88	0.004	.6956377	.9331798
vetsci	.4057399	.126613	-2.89	0.004	.2201032	.7479439
agrisci	.4903089	.0914747	-3.82	0.000	.3401458	.7067641
physsci	.3464236	.028379	-12.94	0.000	.2950374	.4067596
mathsci	.1264487	.0131685	-19.86	0.000	.1031025	.1550814
compsci	.524632	.0458001	-7.39	0.000	.4421253	.6225356
engtech	.4290742	.0428249	-8.48	0.000	.3528385	.5217816
archbuild	.6020389	.0939019	-3.25	0.001	.4434657	.8173143
law	.9802258	.1038785	-0.19	0.851	.7963809	1.206511
busadmin	.7242	.0553652	-4.22	0.000	.6234247	.8412654
masscomm	1.934107	.2780667	4.59	0.000	1.45916	2.563644
lang	1.389176	.0534393	8.54	0.000	1.288288	1.497965
histphilo	1.592767	.1902635	3.90	0.000	1.260295	2.012947
artdesign	.8892922	.0670106	-1.56	0.119	.7671921	1.030825
educ	.6851643	.0398954	-6.49	0.000	.6112677	.7679943
combined	1.134305	.1197452	1.19	0.233	.9222977	1.395047
russhei	1.111944	.0565563	2.09	0.037	1.006442	1.228506
athomeft	1.023875	.0479213	0.50	0.614	.93413	1.122242
logage	83.55695	9.462193	39.08	0.000	66.92532	104.3217
alevelmix	.7936776	.0255208	-7.19	0.000	.7452014	.8453072
nonalevel	.8331114	.0382566	-3.98	0.000	.761405	.9115708
2						
gender	1.460017	.0266896	20.70	0.000	1.408633	1.513277
totalt~fby10	1.061289	.0008818	71.59	0.000	1.059562	1.063019
disabled	.9087724	.0276915	-3.14	0.002	.856087	.9647001
otherwhite~h	.9406713	.0659336	-0.87	0.383	.8199273	1.079196
blackcareth	.6038396	.0525064	-5.80	0.000	.5092207	.7160398
blackafreth	.5662486	.0396528	-8.12	0.000	.4936282	.6495527
otherblack~h	.794723	.1471667	-1.24	0.215	.5528285	1.14246
indianeth	.7532382	.0276392	-7.72	0.000	.7009685	.8094057
pakistaneth	.7638234	.0424348	-4.85	0.000	.6850208	.8516911
banglaeth	.8324394	.0756654	-2.02	0.044	.6965973	.9947717
chineth	.5609069	.0384146	-8.44	0.000	.4904501	.6414852
otherasian~h	.7940324	.0590554	-3.10	0.002	.6863267	.9186405

mixedeth	.9416231	.0519983	-1.09	0.276	.8450299	1.049258
othereth	.8577698	.0827088	-1.59	0.112	.7100603	1.036206
notknowneth	.8477365	.0601601	-2.33	0.020	.7376577	.974242
rank	1.017524	.0031734	5.57	0.000	1.011324	1.023763
medden	.7739354	.1126934	-1.76	0.078	.5817823	1.029553
alliedmed	.8596312	.0374034	-3.48	0.001	.7893607	.9361572
biosci	1.030095	.0338482	0.90	0.367	.9658454	1.098619
vetsci	.4057399	.126613	-2.89	0.004	.2201032	.7479439
agrisci	.8629502	.0864692	-1.47	0.141	.7090775	1.050214
physsci	.7398226	.0320752	-6.95	0.000	.6795534	.8054371
mathsci	.4777465	.0314806	-11.21	0.000	.4198638	.543609
compsci	1.159942	.0565562	3.04	0.002	1.054225	1.276259
engtech	.7984891	.0461537	-3.89	0.000	.7129655	.8942717
archbuild	.7415625	.0558196	-3.97	0.000	.6398457	.8594491
law	.7938492	.0322952	-5.67	0.000	.7330095	.8597386
busadmin	.9682806	.0357228	-0.87	0.382	.9007367	1.04089
masscomm	1.420128	.0682261	7.30	0.000	1.29251	1.560347
lang	1.389176	.0534393	8.54	0.000	1.288288	1.497965
histphilo	1.487	.0611684	9.65	0.000	1.371817	1.611853
artdesign	1.362358	.047032	8.96	0.000	1.273226	1.457729
educ	.6851643	.0398954	-6.49	0.000	.6112677	.7679943
combined	1.134305	.1197452	1.19	0.233	.9222977	1.395047
russhei	1.177409	.0258075	7.45	0.000	1.127899	1.229093
athomeft	.9875422	.0225864	-0.55	0.584	.9442512	1.032818
logage	83.55695	9.462193	39.08	0.000	66.92532	104.3217
alevelmix	.7936776	.0255208	-7.19	0.000	.7452014	.8453072
nonalevel	.8331114	.0382566	-3.98	0.000	.761405	.9115708
3						
gender	1.015625	.0285903	0.55	0.582	.9611066	1.073235
totalt~fby10	1.061289	.0008818	71.59	0.000	1.059562	1.063019
disabled	.9087724	.0276915	-3.14	0.002	.856087	.9647001
otherwhite~h	.9406713	.0659336	-0.87	0.383	.8199273	1.079196
blackcareth	.6038396	.0525064	-5.80	0.000	.5092207	.7160398
blackafreth	.5662486	.0396528	-8.12	0.000	.4936282	.6495527
otherblack~h	.794723	.1471667	-1.24	0.215	.5528285	1.14246
indianeth	.7532382	.0276392	-7.72	0.000	.7009685	.8094057
pakistaneth	.7638234	.0424348	-4.85	0.000	.6850208	.8516911
banglaeth	.8324394	.0756654	-2.02	0.044	.6965973	.9947717
chineth	.5609069	.0384146	-8.44	0.000	.4904501	.6414852
otherasian~h	.7940324	.0590554	-3.10	0.002	.6863267	.9186405
mixedeth	.9416231	.0519983	-1.09	0.276	.8450299	1.049258
othereth	.8577698	.0827088	-1.59	0.112	.7100603	1.036206
notknowneth	.8477365	.0601601	-2.33	0.020	.7376577	.974242
rank	1.017524	.0031734	5.57	0.000	1.011324	1.023763
medden	.8441411	.1230152	-1.16	0.245	.6344114	1.123205
alliedmed	1.367161	.0866062	4.94	0.000	1.207531	1.547893
biosci	1.317423	.0636474	5.71	0.000	1.1984	1.448266
vetsci	.4057399	.126613	-2.89	0.004	.2201032	.7479439
agrisci	1.787589	.2800491	3.71	0.000	1.31497	2.430074
physsci	1.35111	.0891243	4.56	0.000	1.18725	1.537585
mathsci	2.39063	.1826897	11.40	0.000	2.05809	2.776901
compsci	2.818942	.1882218	15.52	0.000	2.473154	3.213077
engtech	2.135492	.1889006	8.58	0.000	1.795571	2.539763
archbuild	1.211549	.1482843	1.57	0.117	.9531485	1.540003
law	.4141265	.0328148	-11.13	0.000	.3545561	.4837056
busadmin	1.160855	.0713327	2.43	0.015	1.029136	1.309432
masscomm	1.106814	.0893834	1.26	0.209	.9447871	1.296628
lang	1.389176	.0534393	8.54	0.000	1.288288	1.497965
histphilo	1.202461	.0664107	3.34	0.001	1.079096	1.339929
artdesign	1.953336	.096275	13.58	0.000	1.773468	2.151447
educ	.6851643	.0398954	-6.49	0.000	.6112677	.7679943
combined	1.134305	.1197452	1.19	0.233	.9222977	1.395047

russhei	.9624103	.0286244	-1.29	0.198	.9079114	1.020181
athomeft	1.177031	.0443095	4.33	0.000	1.093312	1.26716
logage	83.55695	9.462193	39.08	0.000	66.92532	104.3217
alevelmix	.7936776	.0255208	-7.19	0.000	.7452014	.8453072
nonalevel	.8331114	.0382566	-3.98	0.000	.761405	.9115708