Analysis of student characteristics and attainment outcomes at the University of Cambridge

Executive summary

This report provides a summary of the results of the quantitative analysis of the attainment outcomes and the links between various student characteristics and such outcomes, with a particular focus on ethnicity and disability.

The analysis is part of the wider programme of action undertaken by the University of Cambridge to gain better understanding of the potentially causes for previously identified gaps in attainment, which is a first step towards designing best-placed interventions aimed at eliminating such gaps. It aims to help The University to fulfil its commitment specified in 2020-21 to 2024-25 Access and Participation Plan (APP) to complete an investigation of the reasons behind the attainment gaps before determining appropriate reduction targets.

The summary of key findings from a series of statistical models based on individualised datasets covering periods from 2011-12 to 2018-19 are as follows:

1. Strongest predictors of the attainment in the final year of study based on the value of the coefficients of determination (R-squared) of univariate models – i.e. models where the influence of each characteristic on attainment outcomes was tested independently of all other factors – in the order of their predictive strength are:

    - results of first year examinations;
    - course/Tripos.

2. For all three types of outcomes considered in this analysis, the impact on attainment of being from Black or Asian ethnic groups remains statistically significant in the models that controlled for all other factors, including prior attainment. When comparing these two affected groups, size of the gaps in per cent point difference is higher for students from Black ethnic groups than those for students from Asian ethnic groups.

3. The impact of these two ethnic groups remains significant in the models for first class and 'good honours' outcomes in the final year even when the outcomes of the first year examination results – strongest predictor of the final year outcome – are controlled for, although the gap does reduce significantly.

4. None of the models retain any of the two-way interaction terms between Black ethnicity and other factors, such as POLAR, IMD or gender as significant.

5. Thus, the attainment gap for students in Black and Asian ethnic groups remains largely unexplained by the factors included in this analysis. It should be noted that although we were able to control for a number of characteristics, there is still a range of factors that can impact the performance, such as parental attributes (level of education and income), forms of study behaviour, peer-group integration, participation in extra-curricular activities, as well as forms and types of assessments. Further qualitative and

student-led projects are starting to investigate these aspects to assess their influence on the attainment.

6. When the effect of disability on the outcomes is examined independently from all other factors for examinations in all years, being in mental health disability group shows significant influence on obtaining the first class result. However for 'good honours', the outcomes are significantly different for all disability groups with the exception of sensory, medical or physical group, with the gap being highest for social or communication impairment group.

7. In models of the final year outcomes that control for all other factors, the significant effect of mental health group on the first class results obtained is removed, indicating that factors other than belonging to this group is causing this outcome for the first class results.

8. When all other factors, including prior attainment and first year results, are controlled for, the impact of being in any of the disability groups with the exception of sensory, medical or physical group on obtaining 'good honours' and average per cent mark remains significant.

9. These findings suggest that attainment gap for most disability groups cannot be explained by other factors and the outcomes are influenced by the fact of belonging to a disability group itself or some other factors not considered in this analysis.

Introduction

In spring 2019, as part of the development of the 2020-21 to 2024-25 Access and Participation Plan (APP), the University of Cambridge conducted a self-assessment of data on gaps in continuation and attainment between different student groups. The exercise demonstrated the absence of any persistent trends in gaps in attainment due to participation (POLAR4) or deprivation (IMD) markers and identified sustained gaps in attainment between ethnic and disability groups as the main focus of work to ensure successful outcomes for all students.

The self-assessment exercise was based on interrogation of the descriptive statistics and analysis of time series trends. An interactive dashboard of summary statistics and key APP indicators and gap was created and is now available to members of staff and students within the University through Tableau Server, with annual updates incorporating new releases of data by the Office for Students (OfS).

The monitoring work on gaps in attainment continues to be delivered through the annual publication of the undergraduate examination results statistics, which now includes standard reports on gaps between different student groups.

A number of departments and Faculties over the years performed inferential studies of the attainment gaps, particularly with regards to the gender gaps in obtaining the first class results. However, little quantitative analysis of potential causes was available centrally.

As in the 2020-21 to 2024-25 APP the University made a commitment to investigate the reasons for these gaps before determining a target and putting in place procedures for reducing these, it was considered necessary to carry out examinations of factors influencing attainment. This report is a part of the suite of a number of quantitative and qualitative studies, which were undertaken through the 2020-21 academic year to fulfil this commitment.

Methodology

Data sources

Two separate data source was available to perform the analysis:

1. Individualised APP data source as supplied by the Office for Students (OfS) in March 2019. This data source derives from the Higher Education Statistics Agency (HESA) Student record data return submitted by the University in years 2013-14 through to 2017-18.

2. Individualised undergraduate examination results data set covering the period from 2011-12 to 2018-19 and containing annual outcomes for all University of Cambridge undergraduate-level examinations.

Preliminary analysis was carried out using the OfS data source as it held an advantage of using the same data as in the self-assessment exercise. However, for the final modelling the analysis shifted to using the internal data source for the following reasons:

1. Unlike most Higher Education institutions in the UK, the University of Cambridge does not assign a class to an overall degree; instead individual annual examinations (Tripos Parts) are classed. The outcomes of first degrees reported in the student return to HESA represent the results of the student's final year of study. As a result, degree outcomes for students on Engineering, Computer Sciences and Mathematics Triposes aiming for the Integrated Masters are reported as unclassed, because the relevant Tripos Part examinations in fourth year are awarded on the pass/fail basis. Consequently, an important sub-set of students' results are not included in the OfS data source. The internal data source included third year classed results for these students and thus were available for analysis.

2. It was deemed important to include the year of course as a contributing factor and the analysis confirmed that outcome in the first year examinations was the strongest predictor for final year examinations.

3. The internal University data set contains overall per cent mark and rank (for 2018-19 only) as well as class result, providing extra dependent variables.

Although the final models was based on the internal data source, they were also applied to the OfS APP data source to ensure that there was no differences in findings based on the use of a particular data source.

Data from all available academic years were combined to ensure that the samples of students in particular students groups are large enough to carry out the analysis.

Research questions

This report aims to address the following research questions:

1. For undergraduate students in classed examinations, what factor or intersection of factors constitute best predictors of achieving first class result?
2. For undergraduate students in classed examinations, what factor or intersection of factors constitute best predictors of achieving 'good honours' result?
3. For undergraduate students, what factor or intersection of factors constitute best predictors of their overall per cent mark?
4. Does the gap in attainment between different ethnic groups measured on the basis of any of the three outcomes (first class, good honours, per cent mark) remain significant when other predictor factors are controlled for?
5. Does the gap in attainment between different disability groups measured on the basis of any of the three outcomes (first class, good honours, per cent mark) remain significant when other predictor factors are controlled for?

Classed examinations refers to the University of Cambridge annual Tripos Part results rather than outcomes of individual exams (papers). As noted earlier, in most cases the outcomes of these annual examinations are assigned a class, following the standard UK degree classification system, e.g. the award of a first, second or third class, with the second class normally separated into an upper division and a lower division. However, some examinations have unclassed outcomes, i.e. they are awarded on the pass/fail basis. This is similar again to the UK-wide sector practice of not classifying degrees in Medicine, Dentistry and Veterinary Science.

Students who receive non-classed results should not be regarded as having failed to obtain any specific class of honours, but logically they should not be taken into account in calculating the likelihood of obtaining first class or 'good honours'.

Factors considered

The students' characteristics contained in data sources, which were used as independent variables for the analysis include the following:

- Gender
- Ethnicity group
- Disability group
- Age group (young vs mature)
- Secondary school type
- Course of study (Tripos)
- Month of birth
- POLAR4 quintile for UK-domiciled students only
- IMD quintile for UK-domiciled students only
- Previous attainment as measured by UCAS tariff for students with tariff-bearing qualifications on entry
- Previous attainment as measured by A-level score. For this factor, A-level grades were converted to numeric scores using UCAS mapping whereby A* = 6, A = 5 etc.
- Number of A-levels obtained
- Examination results in the first year of study

The above characteristics were identified based on the assumption that none of them should influence the attainment outcome but where the evidence from the University's own data or UK sector data shows that outcomes for the groups of students nonetheless differ. The exception to this rule were two variables measuring previous attainment where there is a well-documented link between previous and subsequent attainment. The inclusion of these factors was necessary to ensure that we were not misinterpreting their influence as that being caused by another one of the variables.

Earlier versions of the model included further variables, such as student's country of permanent address prior to the start of the course (domicile) but they did not show any consistent links with the outcomes and the distribution of students within the groups was heavily unbalanced. As a result, it was excluded from the subsequent models, the results of which are reported in this paper.

The tables presented in Annex 1 below provide a breakdown of the numbers of students in each group by characteristic and the percentage of the total population that this group represents in the internal examination results data set for all academic years combined.

It should be further noted that:

- In analysis involving the gender variable, students with gender recorded as 'other' were excluded due to very small numbers (0.05% of total population across combined academic years).
- In analysis involving ethnicity, the groups 'information refused' and 'unknown' were not considered as separate groups. Otherwise, for the purposes of this analysis the five groups used were consistent with OfS APP analysis, so that students reporting their ethnicity as Chinese were included in the Asian ethnicity group, and those reporting their ethnicity as Gypsy were included in the White ethnicity group. A more detailed breakdown of ethnic groups led to sample groups with very small numbers and thus was not pursued, however, it is important to note that the results might therefore mask differences within particular ethnic groups.
- Due to small numbers in some POLAR4 and IMD quintiles, the quintiles were combined into two groups: quintiles 1-2 and quintiles 3-5.
- In analysis of attainment gap between students in Black and White ethnicity groups, two-way interactions between ethnicity and other characteristics were included, but it should be noted that some of these interactions resulted in samples with small numbers of students.

The models used three different dependent variables as measures of attainment outcome:

1. binary flag indicating whether or not first class outcome was obtained;
2. binary flag indicating whether or not 'good honours' (first or 2:1 class) outcome was obtained;
3. Overall per cent mark.

Analyses conducted

The models testing the first two dependent variables used a binary logistic regression analysis, as it examines the probability of a binary outcome (either a student obtains a first class/good

honours result or not). The analysis of the interval per cent mark variable used a linear regression analysis.

As a first step, a number of univariate regression models were run, each of which examined a single factor from the list identified above separately. This enabled the estimation of the goodness of fit of each univariate model based on R-squared statistics, which indicates the proportion of the variance in the dependent variable that the independent variables explains.

Then multivariate models were run, which included all of the factors as multiple independent variables. In all cases, for each group of students defined by every possible combination of the characteristics specified in section 2 above, the statistical models calculated the probability of a given outcome. A stepwise selection method was then used with an entry criterion of $\alpha=0.05$ and a stay criterion of $\alpha=0.05$ to retain only variables where there is a statistically significant relationship between each characteristic and the outcome.

It should be noted that for groups which contain small numbers of students, the outcomes of these students would not necessarily reflect the behaviour of a larger group with the same characteristic.

This methodology is in line with OfS's research on the associations between characteristics of students as published on 26 September 2019 here: https://www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students/.

Findings

Results of univariate model analysis

A total of 33 univariate models were run (11 factors x 3 outcomes). As the examination results in the first year of study was investigated as one of the factors, it is important to note that all three types of outcomes relate to the final years of study. In case of Integrated Masters this was the third year of study to ensure the availability of classed results across all courses.

The Table 1 below details the R-squared measure for each model, which indicates the predictive ability of a particular variable – the higher the value of R-squared, the more variance in the outcome (dependent variable) it explains. The factors are then ranked on the basis of the R-squared values.

Table1

Summary of the goodness of fit of different factors based on univariate models for three different outcomes

Predictors	**Univariate model**					
	Good Honours		**First class**		**Per cent mark**	
	R-squared	**rank**	**R-squared**	**rank**	**R-squared**	**rank**
Year 1 outcome	0.110	1	0.053	1	0.304	1
Course	0.100	2	0.013	2	0.056	2
Gender	0.019	3	0.001	9	0.000	10
Ethnicity	0.011	4	0.003	5	0.008	3
Disability	0.007	5	0.001	10	0.005	5
Secondary school type	0.006	6	0.004	4	0.003	7
Number of A levels	0.006	7	0.000	11	0.000	11
A-level score	0.002	8	0.003	6	0.002	9
Month of birth	0.002	9	0.001	8	0.005	4
Age group	0.001	10	0.002	7	0.003	8
UCAS tariff	0.000	11	0.005	3	0.004	6

The results show that taken individually, all of the factors remain fairly poor predictors of the outcome. For example, the top ranked predictor based on R-squared value for the first class outcome variable only explains 5.3% of the variance in that outcome.

Across all three outcomes, the results obtained in the first year of study and the course of study (Tripos) remain the two strongest predictors in that order.

It is interesting to note that of all student characteristics factors, ethnicity grouping is the more consistent with regards to being ranked higher on the basis of R-squared statistics. However, its predictive ability remains very low, with 1.1% of variance explained for the good honours outcome.

For two out of three outcomes, previous attainment based on UCAS tariff performs slightly better than A-level score as a predictor, however both are very poor and the slightly better performance of the tariff score might be explained by the fact that it is available for a larger number of students in the data set (e.g. students with International Baccalaureate or Pre-U as previous qualifications).

In addition to examining the strength of various factors as predictors, the models examined which particular categories within each model were found to be statistically significant. The Table 2 below summarises the outcomes – where a particular factor for a model was not statistically significant, it is not listed in the table. It is important to note that the statistically significant result for a particular category does not indicate the importance or causal effect on the outcomes – as highlighted above, the predictive powers of all the factors on the outcome remain very low. The significant results are reported for completeness of the models' presentation.

Table 2

Summary of the statistically significant results from univariate regression models

Dependent variable (outcome)	**Independent variable (factor)**	**Statistically significant results**
Good Honours	Course	A number of courses had **lower** probability of obtaining the outcome.
	Age group	Mature student group had **lower** probability of obtaining the outcome (coefficient = -0.47, SE = 0.18, z = -2.58, p< 0.001).but only at a significance level of α = 0.05 and with unbalance sample sizes for the binary factor. This results should be seen in the context of combined academic year dataset where for years up to 2014-15, Theology for Ministry course, which specialises in admitting mature students, did not separate second class into divisions. Undivided second class result is not included within 'good honours' definition.
	Gender	Male student group had **lower** probability of obtaining the outcome (coefficient = -0.78, SE = 0.07, z = -11.97, p< 0.001). There is likely to be an interaction with the results for the course of study factor, as all courses with significantly lower probability of obtaining results have a higher proportion of male students.
	Month of birth	Several months of birth had **lower** probability of obtaining the outcome, both at a significance level of α = 0.05: • June (coefficient = -0.33, SE = 0.15, z = -2.16, p < 0.05); • August (coefficient = -0.34, SE = 0.15, z = -2.25, p < 0.05); • December (coefficient = -0.33, SE = 0.15, z = -2.40, p < 0.05).
	First year outcome	The students who obtained good honours in the first year examinations had **higher** probability of achieving good honours in their final year (coefficient = 1.85, SE = 0.07, z = 27.71, p < 0.001)
	Secondary school type	Three secondary school types had **lower** probability of obtaining the outcome: • Comprehensive (coefficient = -0.38, SE = 0.08, z = -4.50, p < 0.001); • State grammar (coefficient = -0.32, SE = 0.08, z = -3.90, p < 0.001); • State other (coefficient = -0.50, SE = 0.09, z = -5.90, p < 0.001) There is likely to be an interaction with the results for the course of study factor, as science courses with lower probability of obtaining good honours have a higher proportion of students from state schools.
	Ethnicity	For a number of ethnic groups the probability of obtaining the outcome was significantly **lower**: • Asian (coefficient = -0.63, SE = 0.09, z = -7.28, p < 0.001); • Black (coefficient = -1.20, SE = 0.19, z = -6.46, p < 0.001); • Mixed (coefficient = -0.26, SE = 0.13, z = -1.97, p < 0.05); • Other (coefficient = -0.65, SE = 0.25, z = -5.22, p < 0.01).
	Disability	For a number of disability groups the probability of obtaining the outcome was significantly **lower**: • Cognitive or learning difficulties (coefficient = -0.31, SE = 0.13, z = -2.38, p < 0.05); • Mental health condition (coefficient = -0.62, SE = 0.15, z = -4.01, p < 0.001);

		• Other or multiple impairments (coefficient = -0.77, SE = 0.18, z = -4.22, p < 0.001); • Social or communication impairment (coefficient = -1.24, SE = 0.23, z = -5.36, p < 0.001).
First class	Course	Three courses had **higher** probability of obtaining the outcome, whilst one had had **lower** probability of obtaining the outcome but only at a significance level of α = 0.05. It should be further noted that this course has very low number of students.
	Age group	Mature student group had **lower** probability of obtaining the outcome (coefficient = -0.82, SE = 0.17, z = -4.79, p < 0.001).
	Gender	Male student group had **higher** probability of obtaining the outcome (coefficient = 0.14, SE = 0.04, z = 3.64, p < 0.001).
	First year outcome	The students who obtained first class in the first year examinations had **higher** probability of achieving first class in their final year (coefficient = 1.29, SE = 0.05, z = 26.14, p < 0.001).
	Secondary school type	Three secondary school types had **lower** probability of obtaining the outcome: • Comprehensive (coefficient = -0.29, SE = 0.05, z = -5.32, p < 0.001); • State grammar (coefficient = -0.28, SE = 0.05, z = -5.20, p < 0.001); • State other (coefficient = -0.28, SE = 0.06, z = -4.92, p < 0.001).
	Ethnicity	For a number of ethnic groups the probability of obtaining the outcome was **lower**: • Asian (coefficient = -0.29, SE = 0.07, z = -4.14, p < 0.001); • Black (coefficient = -1.13, SE = 0.23, z = -4.88, p < 0.001).
	Disability	Students in mental health condition group had **lower** probability of obtaining the outcome (coefficient = -0.34, SE = 0.13, z = -2.60, p < 0.01).
Per cent mark	Course	10 courses had **higher** per cent marks whilst two courses had had **lower** per cent marks.
	Age group	Mature student group had **lower** per cent marks (coefficient = -2.59, SE = 0.52, t = -5.72, p < 0.001).
	Gender	Male students had **lower** per cent marks (coefficient = -0.28, SE = 0.14, t = -2.00, p < 0.05) but only at a significance level of α = 0.05.
	First year outcome	The students with a higher first year examinations' per cent mark had **higher** final year per cent mark (coefficient = 0.53, SE = 0.01, t = 71.47, p < 0.001).
	Secondary school type	Three secondary school types had **lower** per cent marks: • Comprehensive (coefficient = -0.75, SE = 0.19, t = -3.96, p < 0.001); • State grammar (coefficient = -0.82, SE = 0.18, t = -4.46, p < 0.001); • State other (coefficient = -0.95, SE = 0.20, t = -4.74, p < 0.001).
	Ethnicity	Three ethnic groups had **lower** per cent marks: • Asian (coefficient = -1.73, SE = 0.23, t = -7.51, p < 0.001);

		• Black (coefficient = -3.82, SE = 0.60, t = -6.34, p < 0.001); • Mixed (coefficient = -0.79, SE = 0.31, t = -2.59, p < 0.01).
	Disability	Three disability groups had **lower** per cent marks: • Cognitive and learning difficulties (coefficient = -0.76, SE = 0.32, t = -2.37, p < 0.05); • Mental health condition (coefficient = -2.91, SE = 0.43, t = -6.78, p < 0.001); • Other or multiple impairments (coefficient = -1.98, SE = 0.54, t = -3.69, p < 0.001).

Table 3

Summary of the statistically significant results from multivariate regression models

Model (outcome)	**Factor**	**Statistically significant results**
Good honours (overall model McFadden R-squared = 0.23)	Age group	Mature student group had **lower** probability of obtaining the outcome (coefficient = 1.53, SE = 2.45, z = -3.07, p< 0.001).
	Course	The probability of the outcome remained **lower** for three courses, whilst for five courses had a **higher** probability of good honours results when all other factors are also controlled for.
	Gender	Male student group had a **lower** probability of obtaining the outcome (coefficient = -0.53, SE = 7.86, z = -6.70, p< 0.001).
	Ethnicity	The probability of the outcome remained **lower** for the three ethnic groups: • Asian (coefficient = -0.48, SE = 1.04, z = -4.64, p< 0.001); • Black (coefficient = -1.21, SE = 2.36, z = -5.14, p< 0.001); • Mixed (coefficient = -0.30, SE = 1.49, z = -2.02, p< 0.05).
	Disability	The probability of the outcome remained **lower** for the four disability groups: • Cognitive or learning difficulties (coefficient = -0.45, SE = 1.51, z = -2.30, p< 0.01); • Mental health condition (coefficient = -1.04, SE = 1.90, z = -5.46, p< 0.001); • Other or multiple impairments (coefficient = -0.92, SE = 2.22, z = -4.12, p< 0.001); • Social or communication impairment (coefficient = -1.05, SE = 2.87, z = -3.65, p< 0.001).
	Secondary school type	The results for the comprehensive school type was no longer statistically significant in the multivariate model and the results for state grammar type was only significant at a reduced to significance level of α = 0.05 (coefficient = -0.21, SE = 9.58, z = -2.18, p< 0.05). However, the results for state other category remained statistically significant (coefficient = -0.33, SE = 9.85, z = -3.37, p< 0.001).
	Month of birth	The probability of obtaining the outcome remained **lower** for one month category – June, reduced to significance level of α = 0.05 (coefficient = -0.34, SE = 1.74, z = -1.98, p< 0.05).
	First year outcome	The results for the outcome in the first year examinations remained statistically significant (coefficient = 1.72, SE = 9.23, z = 18.64, p< 0.001).

First class (overall model McFadden R-squared = 0.15)	Course	All three courses that had **higher** probability of first class results in the univariate model, retained their significant result when all other factors are also controlled for, and one further course became statistically significant at α = 0.05 level. There was a further statistically significant results for five courses showing a **lower** probability of first class results.
	Ethnicity	The probability of the outcome remained **lower** for the two ethnicity groups: • Asian (coefficient = -0.23, SE = 0.08, z = -2.77, p< 0.01); • Black (coefficient = -0.91, SE = 0.27, z = -3.41, p< 0.001);
	Secondary school type	The probability of the outcome remained **lower** for the three secondary school types: • Comprehensive (coefficient = -0.20, SE = 0.06, z = -3.13, p< 0.01); • State grammar (coefficient = -0.30, SE = 0.06, z = -4.81, p< 0.001); • State other (coefficient = -0.24, SE = 0.07, z = -3.57, p< 0.001)
	First year outcome	The results for the outcome in the first year examinations remained statistically significant (coefficient = 1.64, SE = 0.06, z = 28.20, p< 0.001)
	A-level score	A-level score became a statistically significant factor (coefficient = 0.02, SE = 0.01, z = 4.77, p< 0.001)
Per cent mark (overall model multiple R-squared = 0.09, F = 17.47 on 58 and 10,764 DF)	Age group	Mature student group had **lower** per cent mark (coefficient = -2.57, SE = 0.78, t = -3.30, p< 0.001).
	Course	All courses that had **higher** per cent scores in the univariate model also had statistically significant results in the multivariate model.
	Ethnicity	The per cent mark remained **lower** for the two ethnicity groups: • Asian (estimate = -1.41, SE = 0.24, t = -5.94, p< 0.001); • Black (estimate = -3.30, SE = 0.62, t = -5.35, p< 0.001).
	Disability	The per cent mark remained **lower** for the four disability groups: • Cognitive or learning difficulties (estimate = -0.88, SE = 0.33, t = -2.67, p< 0.01); • Mental health condition (estimate = -3.00, SE = 0.44, t = -6.88, p< 0.001); • Other or multiple impairments (estimate = -2.20, SE = 0.55, t = -0.40, p< 0.001); • Social or communication impairment (estimate = -4.73, SE = 0.77, t = -6.15, p< 0.001).
	Secondary school type	The per cent mark remained **lower** for the three secondary school types: • Comprehensive (estimate = -0.70, SE = 0.19, t = -3.63, p< 0.001); • State grammar (estimate = -0.98, SE = 0.19, t = -5.22, p< 0.001); • State other (estimate = -0.87, SE = 0.20, t = -4.32, p< 0.001)
	A-level score	A-level score became a statistically significant factor (estimate = 0.18, SE = 0.02, t = 11.26, p< 0.001)

Results of multivariate model analysis

Three multivariate models were run – one for each type of attainment outcomes (dependent variables). The models included all of the factors described in the methodology section.

First of all, it should be noted that all three models had low R-squared values, indicating that the combination of the quantitative factors are not strong predictors of outcomes and a large proportion of the variance in all three types of outcomes are not explained by the variables considered in the analysis.

Table 3 above lists the variables that were found to have statistically significant relationship between that particular characteristic and the type of the outcome in each of the three models.

For categories that remain statistically significant within the multivariate models that control for all other factors, the indication is that the differences cannot be fully explained by the interplay with other variables, so that the underlying causes of the significant difference in outcomes remain unexplained.

With respect of the area of interest related to ethnicity and disability, it should be noted that the impact on attainment of being from Black or Asian ethnic groups remained statistically significant in all multivariate models that controlled for all other factors, including prior attainment, even though the size of the gaps were reduced. For disability, the significant effects of any group was eliminated in the multivariate model of the first class outcome but for the other two types of attainment outcome, their effects remained and were highest for the mental health condition group.

Conclusions

Both univariate and multivariate modelling of a number of quantitative factors indicate a relatively low predictive strength of these factors, both as independent predictors and as a combined set. It is not possible to numerically capture and model all characteristics of an individual and use them in a predictive model. It can be hypothesised that a number of attributes related to teaching and learning approaches, such as study and revision behaviours, levels of self-confidence, peer-group interactions, content of education activities and assessment types, which are intrinsically harder to measure quantitatively and which thus remained outside of the scope of this investigation, can be influencing the outcomes.

Of all factors studied, the attainment in the first year of the course was consistently shown to be the best predictor amongst the rest, which is not surprising as previous attainment is a known indicator of future academic performance.

Course of study was also identified as a significant predictor but its influence was found to be different depending on which type of outcome was considered. For example, a number of courses had significantly different average per cent mark but no significant difference in classed outcomes. This reflects the potentially different practices of converting marks obtained to classes and it can be recommended that further work is carried out to document such practices for each course in one place. This recommendation might further help with analysis of classification outcomes inflation (often referred to as grade inflation).

In terms of investigating the attainment gaps for different ethnicity and disability groups, this analysis confirmed the conclusions of the APP self-assessment that there are significant gaps between students from Black and White ethnicity groups, Asian and White ethnicity group and a number of disability groups, particularly mental health conditions. The fact that the lower outcomes for these groups persisted in the multivariate models where other factors were controlled for indicates that underlying causes of the difference remain partially unexplained by the variables included in this study and that intervention work needs to focus on exploring "softer" less numerical attributes related to teaching and learning practices.

Overall, as the selection of ethnicity and disability amongst the factors was predicated on the assumption that in and of themselves these student characteristics should not be influencing attainment, the outcomes of the analysis suggest that the University should target the activities aimed at removing any unexplained gaps in performance between students of different ethnic or disability groups.

Dr Ekaterina Samoylova and Dr Laura Hall
Academic and Financial Planning and Analysis
April 2020

Count of student attainment outcomes in whole population broken down by student group characteristics

Table A1

Descriptive statistics for course of study

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Natural Sciences	14,905	23.76	14,905	23.76
Modern and Medieval Languages	4,964	7.91	19,869	31.67
Engineering	4,402	7.02	24,271	38.69
Mathematics	3,661	5.84	27,932	44.52
Medicine	3,590	5.72	31,522	50.25
Law	3,115	4.97	34,637	55.21
English	2,896	4.62	37,533	59.83
History	2,888	4.60	40,421	64.43
Economics	2,427	3.87	42,848	68.30
Human, Social and Political Sciences	2,163	3.45	45,011	71.75
Geography	2,078	3.31	47,089	75.06
Classics	1,611	2.57	48,700	77.63
Music	1,359	2.17	50,059	79.80
Computer Science	1,333	2.12	51,392	81.92
Theology, Religion, and Philosophy of Religion	1,041	1.66	52,433	83.58
Veterinary Medicine	926	1.48	53,359	85.06
Chemical Engineering	899	1.43	54,258	86.49
Politics, Psychology and Sociology	878	1.40	55,136	87.89
Philosophy	864	1.38	56,000	89.27
Asian and Middle Eastern Studies	782	1.25	56,782	90.51
Architecture	768	1.22	57,550	91.74
Land Economy	712	1.13	58,262	92.87
Psychological and Behavioural Sciences	694	1.11	58,956	93.98
History of Art	556	0.89	59,512	94.86
Linguistics	536	0.85	60,048	95.72
Archaeology and Anthropology	529	0.84	60,577	96.56
Education	484	0.77	61,061	97.33
Anglo-Saxon, Norse and Celtic	435	0.69	61,496	98.03
Management Studies	358	0.57	61,854	98.60
Classics (4-year course)	229	0.37	62,083	98.96
Theology for Ministry	228	0.36	62,311	99.33
Manufacturing Engineering	218	0.35	62,529	99.67
History and Politics	85	0.14	62,614	99.81
History and Modern Languages	61	0.10	62,675	99.91
Archaeology	59	0.09	62,734	100.00

Table A3

Descriptive statistics for gender

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	33,217	52.95	33,217	52.95
Female	29,517	47.05	62,734	100.00

Table A4

Descriptive statistics for disability

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No declared disability	56,270	89.70	56,270	89.70
Cognitive or learning difficulties	2,559	4.08	58,829	93.78
Mental health condition	1,249	1.99	60,078	95.77
Other or multiple impairments	989	1.58	61,067	97.34
Sensory, medical or physical impairments	888	1.42	61,955	98.76
Social or communication impairment	458	0.73	62,413	99.49
Information refused	307	0.49	62,720	99.98
Not known	14	0.02	62,734	100.00

Table A5

Descriptive statistics for ethnicity

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
White	50,208	80.03	50,208	80.03
Asian	6,528	10.41	56,736	90.44
Mixed	3,247	5.18	59,983	95.61
Information refused	1,269	2.02	61,252	97.64
Black	846	1.35	62,098	98.99
Other	636	1.01	62,734	100.00

Table A6

Descriptive statistics for secondary school type

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Independent	23,878	38.06	23,878	38.06
Comprehensive	13,919	22.19	37,797	60.25
State Grammar	13,490	21.50	51,287	81.75
State Other	9,494	15.13	60,781	96.89
Other	1,506	2.40	62,287	99.29
FE	375	0.60	62,662	99.89
Not known	72	0.11	62,734	100.00

Table A7

Descriptive statistics for month of birth

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
October	6,028	9.61	6,028	9.61
September	5,813	9.27	11,841	18.87
March	5,393	8.60	17,234	27.47
April	5,336	8.51	22,570	35.98
December	5,254	8.38	27,824	44.35
November	5,214	8.31	33,038	52.66
May	5,180	8.26	38,218	60.92
January	5,163	8.23	43,381	69.15
August	4,903	7.82	48,284	76.97
June	4,850	7.73	53,134	84.70
February	4,821	7.68	57,955	92.38
July	4,779	7.62	62,734	100.00

Table A8

Descriptive statistics for age group

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Young	61,208	97.57	61,208	97.57
Mature	1,526	2.43	62,734	100.00

Table A9

Descriptive statistics for number of A-levels

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
4	28,071	44.75	28,071	44.75
3	19,434	30.98	47,505	75.72
5	7,776	12.40	55,281	88.12
No A-levels	4,792	7.64	60,073	95.76
6	1,237	1.97	61,310	97.73
2	746	1.19	62,056	98.92
1	479	0.76	62,535	99.68
7	163	0.26	62,698	99.94
8	31	0.05	62,729	99.99
9	5	0.01	62,734	100.00

Table A10

Descriptive statistics for A-level score

Value	Frequency	Percent	Cumulative Frequency	Cumulative Percent
24	10,276	16.38	10,276	16.38
23	7,177	11.44	17,453	27.82
18	6,417	10.23	23,870	38.05
17	6,377	10.17	30,247	48.21
No A-levels	4,792	7.64	35,039	55.85
22	4,633	7.39	39,672	63.24
16	4,344	6.92	44,016	70.16
20	3,034	4.84	47,050	75.00

21	2,484	3.96	49,534	78.96
15	2,365	3.77	51,899	82.73
29	1,857	2.96	53,756	85.69
30	1,704	2.72	55,460	88.41
28	1,348	2.15	56,808	90.55
27	1,023	1.63	57,831	92.18
25	934	1.49	58,765	93.67
26	621	0.99	59,386	94.66
19	524	0.84	59,910	95.50
12	400	0.64	60,310	96.14
11	286	0.46	60,596	96.59
34	246	0.39	60,842	96.98
5	231	0.37	61,073	97.35
6	230	0.37	61,303	97.72
14	214	0.34	61,517	98.06
33	213	0.34	61,730	98.40
35	212	0.34	61,942	98.74
36	178	0.28	62,120	99.02
10	160	0.26	62,280	99.28
32	95	0.15	62,375	99.43
31	85	0.14	62,460	99.56
9	40	0.06	62,500	99.63
41	36	0.06	62,536	99.68
4	31	0.05	62,567	99.73
13	29	0.05	62,596	99.78
39	24	0.04	62,620	99.82
40	18	0.03	62,638	99.85
42	18	0.03	62,656	99.88
37	16	0.03	62,672	99.90
38	15	0.02	62,687	99.93
45	11	0.02	62,698	99.94
0	8	0.01	62,706	99.96
3	6	0.01	62,712	99.96
47	5	0.01	62,717	99.97
48	4	0.01	62,721	99.98
52	4	0.01	62,725	99.99
1	3	0.00	62,728	99.99
2	2	0.00	62,730	99.99
44	2	0.00	62,732	100.00
7	1	0.00	62,733	100.00
8	1	0.00	62,734	100.00