Department for Education

Teaching Excellence and Student Outcomes Framework: analysis of final award

Research report

October 2017

Contents

List of figures

List of tables

Executive summary

The Government has considered the lessons learned from Year Two of the Teaching Excellence and Student Outcomes Framework[1] (TEF), for which outcomes were published on 22 June 2017. The lessons learned exercise has focused on the practical operation of the TEF, in line with the commitments made earlier this year. The TEF Year Two lessons learned exercise focused on six main areas:

1. Whether the process of application and assessment worked smoothly and effectively;
2. Whether the guidance to providers was clear and understandable;
3. The way in which the metrics were used, in particular the use of significance flags and their role in generating initial hypotheses;
4. The balance of evidence between core metrics and additional evidence;
5. Whether commendations should be introduced for the next round of TEF assessments;
6. The number and names of the different ratings and their initial impact internationally.

The lesson learned report (https://www.gov.uk/government/publications/teaching-excellence-framework-lessons-learned) addresses all six areas and is supported by two key pieces of analytical work that address the weighting of the National Student Survey[2] (NSS) metrics in TEF (https://www.gov.uk/government/publications/teaching-excellence-framework-analysis-of-metrics) and biases by provider and student characteristics (this report). Both feed into area three.

This research annex specifically looks at whether the assessment process was fair by considering how the TEF final award is related to provider and student characteristics. In particular it considers stakeholder[3] concerns that providers in London might be systematically disadvantaged because they do worse on the NSS and that the deprivation measure POLAR[4] (Participation of Local Areas) does not accurately reflect social disadvantage in London.

The analyses in this report examine whether there are relationships between providers, regions, student characteristics and the final awards. Initial analyses comprise of simple tests of association, with further analyses using ordinal logistic regression techniques. The latter allows the impact of each variable to be assessed whilst controlling for the

[1] See http://www.hefce.ac.uk/lt/tef/whatistef/
[2] For more information see http://www.hefce.ac.uk/lt/nss/
[3] See for example https://www.londonhigher.ac.uk/ceo-blog/london-heis-tef/
[4] For more information see http://www.hefce.ac.uk/analysis/yp/POLAR/

effects of others, which is important as characteristics are highly unlikely to operate independently. The analysis has been peer reviewed by Professor Gavin Shaddick, Chair of Data Science and Statistics at the University of Exeter, see Annex B for the full review.

The aim of the analysis is to investigate if there are specific provider, region or student characteristics that are significantly associated with the TEF final award.

Are there specific provider or student characteristics that bias the TEF final award? Is there a regional bias?

Key findings

- There was no statistically significant difference (after adjusting for the effects of other characteristics) between the proportion of bronze, silver or gold awards between the different provider types (Higher Education Institutes (HEIs), Further Education Colleges and Alternative Providers).
- The proportion of bronze awards in HEIs with low entry tariff is higher than in those with high entry tariff but the difference is **not** statistically significant.
- The proportion of bronze awards is higher for providers in London/South East area compared to providers located elsewhere (after adjusting for the effects of other characteristics) but the difference is **not** statistically significant.
- The proportion of gold awards is lower for providers in London/South East area compared to providers located elsewhere (after adjusting for the effects of other characteristics) but the difference is **not** statistically significant.
- There is no evidence that having a higher percentage of students from deprived areas has an adverse effect on getting a gold award.
- For providers with a high percentage of older students (aged over 30 at start of study) who are local students, the proportion of bronze awards is significantly lower than for those who have high percentage of older students who are not local students.
- The student characteristics, ethnicity; gender and disability, are not statistically associated with TEF award type.

Implications of the findings

1. We note that the detailed analyses showed that the TEF outcomes were not affected by the characteristics of students (ethnicity, gender, disability or background), nor the region where a provider was located.

2. Overall, we consider that the flexible approach to addressing region in Year Two has

worked. It has allowed the TEF panel to consider regional issues in a holistic, considered way. There is no evidence (the analysis shows that region is not significantly associated with the final award) that any systematic regional issues have impacted the ratings.

3. We will not therefore make any change to the framework on regional issues. We will, however, make an explicit reference to local students in the guidance to TEF assessors.

4. In the analyses, allowance has been made for the small degree of correlation between final TEF outcome and provider tariff, number of part-time students and proportion of older and local students.

5. We will be making a number of changes to the assessment process for providers with significant proportions of part-time students, whilst working within the overall assessment framework to maintain consistency for both full and part-time providers.

Introduction

The government has introduced the Teaching Excellence and Student Outcomes Framework (TEF) as a way of better informing students' choices about what and where to study for a higher education qualification. The TEF also aims to raise the esteem for teaching and recognise and reward excellent teaching. The TEF is now in its second year of implementation. The TEF Year two final ratings were recently published[5] giving a rating of gold, silver, bronze or provisional to higher education providers that participated. Participation in TEF is voluntary and in total 295 providers participated. Of these 231 applied for TEF assessment resulting in a gold, silver or bronze award and the remaining 64 providers opted for a provisional award as they did not have enough data for a TEF assessment.

The assessment process looks at core metrics, metrics by subgroups such as ethnicity; deprivation; age and gender, contextual data and additional information that is submitted by the provider to allocate a final award. The contextual data provides information on the nature and operating context of a provider and includes information such as size; location and student population. The metrics are also benchmarked using a number of factors to reduce the likelihood of bias, as it ensures that each provider is measured against other similar providers. See the year two specification[6] for full details.

Following the publication of the final year two results, it was agreed that a lessons learned exercise will be conducted to inform the implementation of TEF Year Three. This analysis looks at whether specific provider characteristics have a higher proportion with the final award than others. One of the most frequently cited regional issues is 'the London effect', which asserts that providers in London are systematically disadvantaged because they do worse on the NSS (for assessment and feedback and academic support) and that POLAR does not accurately reflect social disadvantage in London. This is something that requires particular attention as from the raw numbers it appears that London providers received worse ratings in Year Two than most other regions, particularly when non-specialist institutions are removed from the statistics.

It should be noted that the analysis is based on just the 231 (out of 457) providers who took part in TEF and received a final award of gold, silver or bronze. It excludes providers who received a provisional award. The small size of the dataset should be considered when drawing any conclusions from the analysis.

---

[5] Results are published here http://www.hefce.ac.uk/tefoutcomes

[6] TEF Year Two Specification: http://webarchive.nationalarchives.gov.uk/20170517113229/https://www.gov.uk/government/publications/teaching-excellence-framework-year-2-specification

The data and descriptive statistics

The TEF Year Two data contains data on provider characteristics for the 231 providers who received a gold, silver or bronze award. The contextual data relates to undergraduate students at each provider and is an average of three years' data (or less if the provider did not have three years' worth of data). The characteristics that can be used to predict the outcome are from:

- Provider type (HEI, FEC, AP)
- Region where provider is located
- Number of part-time students
- % students in most deprived category (National Index of Multiple Deprivation[7] (IMD)/POLAR)
- % students who are Black, Asian or other
- % students who are disabled
- % students who are aged over 30[8] at the start of study
- % students who are not male
- % students who are not UK domiciled
- % students who are local[9] students

Before presenting the results of any kind of modelling it is useful to establish the principles that have been followed in the analyses. In the first stage of the analyses, summaries of the data are calculated and associations between variables assessed by examinating differences in the category of award by provider type, student characteristics and region. When comparing proportions, it is important not to have small sample sizes (less than five) in the denominator of the proportion . Where this was not the case, categories have been combined to overcome any sample size issues. Table 1 shows the distribution of the final award by type of provider. Due to the small numbers of Alternative Providers (APs), as shown in Table 1, they have been combined with Further Education Colleges (FECs). It is noted that some APs may be more similar to Higher Education Institutes (HEIs) but their small number means the category to which they are allocated is unlikely to have a meaningful impact on the findings.

---

[7] For the English IMD see https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015, for the Scottish IMD see http://www.gov.scot/Topics/Statistics/SIMD and for the Welsh IMD see http://gov.wales/statistics-and-research/welsh-index-multiple-deprivation/?lang=en

[8] The TEF contextual data contains age at start of study grouped as under 21, 21 to 30 and over 30. A mature student is someone aged over 21 at the start of study.

[9] A student is defined as being local if their home address is within the same Travel to Work Area (TTWA) as their location of study. See here for definition of TTWA https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#travel-to-work-area

Table 1 Distribution of final award for Higher Education Institutes (HEI), Further Education Colleges (FEC) and Alternative Providers (AP)[10]

	HEI	FEC	AP	Total
GOLD	43 (32%)	14 (15%)	2 (33%)	59
SILVER	67 (50%)	46 (51%)	3 (50%)	116
BRONZE	24 (18%)	31 (34%)	1 (17%)	56
Total	134	91	6	231

Table 2 shows the means and standard deviations for the numeric characteristics by award type. For example, the average percentage of students from the most deprived areas (based on IMD/POLAR) in providers who were awarded gold was 28%. For the majority of characteristics, the proportions are similar across the three award types, the notable exception being in the percentage of local students which is lower in providers awarded gold (28%) than for those receiving silver (40%) and bronze (49%).

Table 2 Provider's student characteristics by award type. Mean proportion with standard deviation in brackets

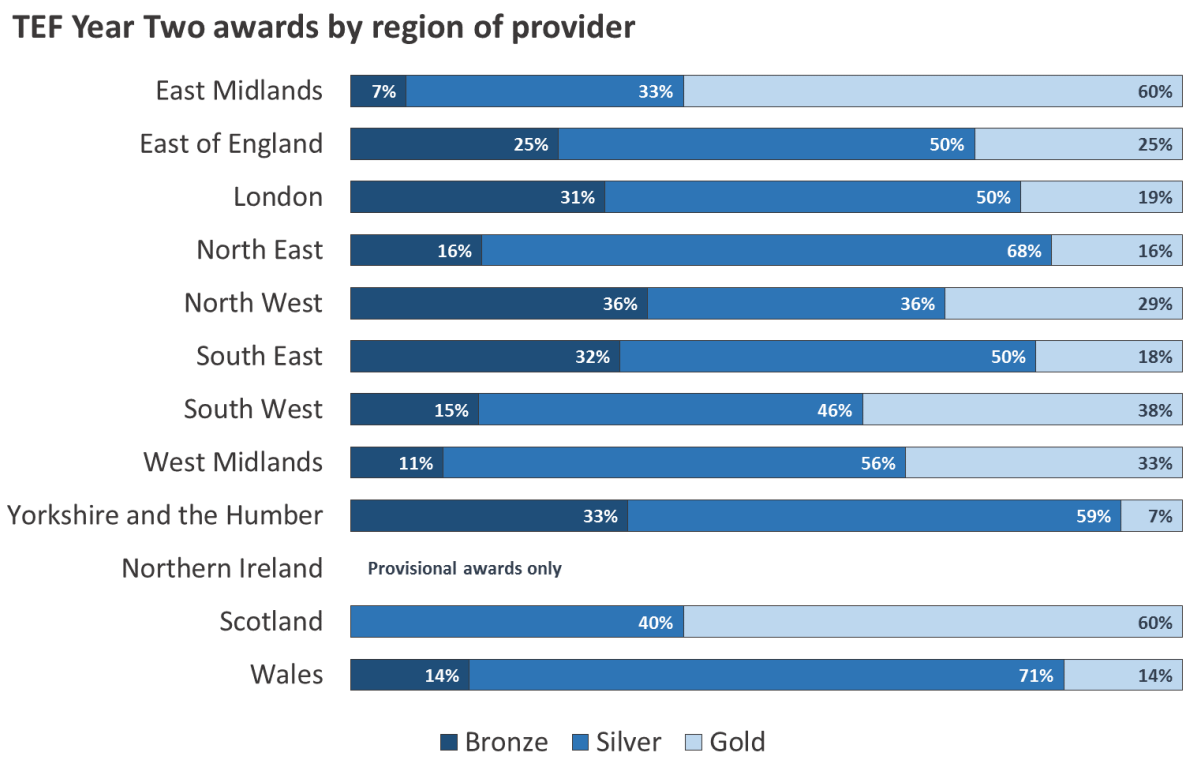
	Mean proportion (and standard deviation) of characteristic for each award type		
Provider information	**Gold**	**Silver**	**Bronze**
Proportion of students who are in the most deprived category (IMD/POLAR)	0.28 (0.13)	0.32 (0.14)	0.32 (0.15)
Proportion of students who are Asian, Black or other BME	0.20 (0.15)	0.21 (0.17)	0.21 (0.20)
Proportion of students who have a disability	0.14 (0.06)	0.13 (0.07)	0.14 (0.05)

[10] This analysis was conducted prior to the outcomes of the appeals process.

Provider information	Mean proportion (and standard deviation) of characteristic for each award type		
	Gold	Silver	Bronze
Proportion of students who are not domiciled in the UK	0.14 (0.12)	0.10 (0.11)	0.08 (0.10)
Proportion of students who are not male	0.55 (0.10)	0.54 (0.11)	0.58 (0.10)
Proportion of students who are aged over 30 at the start of study	0.10 (0.09)	0.15 (0.10)	0.19 (0.11)
Proportion of students who are local students	0.28 (0.23)	0.40 (0.24)	0.49 (0.25)

Figure 1 shows the distribution of the final award by the location of the provider. The wide spread of results across the regions demonstrates that excellent provision can be found nationwide.

Figure 1 Distribution of final award by region where provider is located



Source: HEFCE

When examining possible associations between factors, e.g. student characteristics, that might effect a variable of interest, here the award of gold, silver or bronze, it is important to understand whether the characteristics are correlated with each other. Including highly correlated variables in a statistical model can lead to problems with the estimation of model parameters, the accuracy of the strength of any associations that are found, and the assessment of whether those associations are statistical significant. Table 3 shows the correlation between each of the characteristics.

Table 3 Correlation between numeric characteristics

	% students who are Asian, Black or other BME	% students who are in the most deprived category (IMD/POLAR)	% students who have a disability	% students not UK domiciled	% students who are not male	% students who are aged 30+ at start of study	% students who are local students
% students who are Asian, Black or other BME	1.00	-0.37	-0.31	0.46	-0.05	-0.22	0.03
% students who are in the most deprived category (IMD/POLAR)	-0.37	1.00	-0.07	-0.51	-0.04	0.45	0.40
% students who have a disability	-0.31	-0.07	1.00	-0.20	0.29	-0.07	-0.07
% students not UK domiciled	0.46	-0.51	-0.20	1.00	-0.12	-0.54	-0.37
% students who are not male	-0.05	-0.04	0.29	-0.12	1.00	0.06	-0.09

	% students who are Asian, Black or other BME	% students who are in the most deprived category (IMD/POLAR)	% students who have a disability	% students not UK domiciled	% students who are not male	% students who are aged 30+ at start of study	% students who are local students
% students who are aged 30+ at start of study	-0.22	0.45	-0.07	-0.54	0.06	1.00	0.60
% students who are local students	0.03	0.40	-0.07	-0.37	-0.09	0.60	1.00

As a rule of thumb, characteristics are considered highly correlated with each other if the correlation coefficient is 0.80 or greater. Under this criteria, none of the characteristics considered here are highly correlated. The largest positive correlation is observed between the percentage aged 30+ at start of study and those who are local students and the largest negative correlation between students who are aged 30+ at start of study and students who are not UK domiciled. These might both be expected as older students may often be settled with families and may be more likely to study locally, and for similar reasons students from outside the UK are likely to be younger.

Relationship of provider characteristics with award type

In the first stage of the analysis, associations between characteristics and award type are considered independently, that is, on a characteristic-by-characteristic basis. Here, the assessment of whether any of the characteristics have a relationship with award type, and whether any relationships are statistically significant, is performed using Chi-squared test of independence. This statistical analysis is performed under the assumption that there is no association between award type and a particular characteristic, an assumption holds until evidence is found (in the data) to reject it. The decision whether to reject this assumption is based on the size of the p-value, the probability that award type and a particular characteristic are independent. Technically, the p-value is based on differences between the observed data (proportions in each category) and those which would be expected if there were no relationship between award type and a particular characteristic, that is, they are independent. P-values less than 0.05 are generally considered to indicate that there is evidence that award type and the characteristic in question are not independent, that is, there is a (statistically significant) association between them.

In order to assess the strength of any relationships between award type and continuous variables the data are grouped into quintiles, with each quintile representing an increase in the characteristic.

An illustration of the use of the Chi-squared test to assess the relationship between percentage of students who are Asian, Black or other BME and award type is shown in Table 4. Here, the first quintile consists of providers with the smallest percentages of students who are Asian; Black or other BME with the fifth quintile consisting of providers with the highest percentages. The result of performing the Chi-squared test on this data is the calculation of a test-statistic, the size of which indicates how different the observed data is from what would be expected under the assumption of independence, the degrees of freedom (df), which indicates sample size and determines the threshold which will determine whether statistical significance is achieved, and the p-value which is the measure of (potential) significance.

Table 4 Distribution of the number of providers by final award and proportion of students who are Asian, Black or other BME (categorised into five groups)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
GOLD	10	14	11	15	9
SILVER	23	21	26	22	24
BRONZE	14	11	10	8	13

Chi-squared = 4.91, degrees of freedom = 8, p-value = 0.77

The p-value obtained when performing a Chi-squared test on this data is 0.77. Small values of this probability would lead us to reject the assumption that award type and percentages of students who are Asian, Black or other BME are independent. In this case, the p-value is large (and certainly not less than 0.05) and therefore there is no evidence of a statistically significant association.

Table 5 shows the chi-square values with p-values for all the provider characteristics. Whether the association, between two characteristics, is statistically significant or not is given in the last column.

Table 5 Chi-square value and p-value for all provider characteristics

Characteristic	Chi-square (degrees of freedom)	p-value	Is association with final award statistically significant?
% students who are in the most deprived category (IMD/POLAR)	4.62 (8)	0.798	No
% students who are Asian, Black or other BME	4.91 (8)	0.767	No
Region where provider is located (EM, WM, YH, E/NE, NW, Scot/SW, Wales/SE/London)	5.36 (8)	0.719	No

Characteristic	Chi-square (degrees of freedom)	p-value	Is association with final award statistically significant?
% students who have a disability	5.42 (8)	0.712	No
% students who are not male	8.61 (8)	0.377	No
Region where provider is located (simplified to London/SE or elsewhere)	4.05 (2)	0.132	No
% students not UK domiciled	15.27 (8)	0.054	Yes (as it is only just above 0.05)
Number of part-time students at provider	15.69 (8)	0.047	Yes
Provider type (Specialist HEI/HEI high tariff/HEI medium tariff/HEI low tariff/HEI/FEC or AP)	25.06 (10)	0.005	Yes
Provider type (HEI/FEC or AP)	10.64 (2)	0.005	Yes
Provider type (Specialist HEI/HEI/FEC or AP)	13.35 (4)	0.010	Yes
% students who are local students	27.45 (8)	0.001	Yes
% students who are aged 30+ at start of study	29.77 (8)	< 0.001	Yes

Using this methodology, the following variables are identified as having a statistically significant association with award type:

- % students who are not UK domiciled
- % students who are local students

- % students who are aged over 30 at start of study
- Provider type
- Number of part-time students at provider

However, these analyses only consider the characteristics one-by-one and associations may be driven, or masked, by associations with other characteristics. In order to determine whether these associations are meaningful or may be the result of confounding with other factors, the following analyses allow for the assessment of associations whilst making allowance for the possible effects of other variables.

Model for predicting award type

Where the information contained in a set of variables, e.g. characteristics, cannot be considered to be independent, a regression analysis can be performed to examine the impact of each variable on the response, e.g. the final award, allowing for the effect of the other variables. Logistic regression is used where the response variable is a proportion and there are two possible groups in the response and multinomial regression where there are more than two groups, as here. However, a multinomial regression model does not take into account the ordering of the response variable, e.g. that gold is higher than silver which is higher than bronze. This ordering can be incorporated into the statistical modelling using ordinal logistic regression.

In an ordinal model, the probability of getting lower awards is compared to all awards above that. For example, we can examine the effects of characteristics on getting either silver or gold compared to bronze, and gold compared to bronze or silver. This will allow questions such as ‘what happens to the odds of getting a gold compared to bronze or silver when the percentage of students from a deprived area increases?’ and ‘does it increase or decrease?’ to be answered?

The model results in a set of coefficients for each of the two comparisons: bronze versus gold and silver, and bronze and silver versus gold. For each, the output takes the form of the log odds ratio that represents (on the log-scale) the extent to which an increase (or decrease) in a particular characteristic changes the likelihood of achieving a particular award.

The results presented are from performing these comparisons using an ordinal logistic regression, although it is noted that the pattern of results were the same when multinomial regression was performed in which comparisons are made in the effect of three distinct groups (gold, silver and bronze) but with no allowance for the ordering.

Each model is developed by first considering the provider characteristic that provides the best model fit. In this setting, the provider characteristics are known as covariates. The next stage considers adding a second covariate, again chosen as the one that gives the best model fit (given the first one is already in the model). This process continues until the model fit cannot be improved with the addition of any further covariates. See Annex A for further details on how the final model was chosen. The model building process considers all available variables not just those that were identified as being significantly associated with award type in the initial set of analyses based on the Chi-squared test. The final model may include different variables to those identified earlier as being significantly associated with award type as the regression model considers the effects of all provider characteristics when they are considered together, whereas the earlier analysis looked at each of their effects in isolation.

The final model includes the following characteristics:

- provider type (HEI, FEC/AP)
- Region (London/South East, Elsewhere)
- Number of part-time students
- % students in most deprived category
- % students who are aged over 30 at the start of study
- % students who are local students
- % students who are not domiciled in the UK
- % students who are not male

The rationale for considering London/South East combined versus all other regions is that providers in the London area have argued that they are more likely to get worse results on the TEF than other regions. The year two TEF results (see Figure 1) shows that London and South East have very similar results so one of the aims of the modelling exercise was to investigate if these two regions are adversely affected. A model using just London versus all other regions was also tested and led to the same conclusions.

Due to the small number of APs with full TEF awards (see Table 1) they are considered along with FECs in the model. A more refined version that further breakdown HEIs was also considered, but as with APs there are small sample sizes in some categories, see Table 6.

Table 6 Distribution of TEF final awards by provider type and entry tariff[11]

Provider type	Bronze	Silver	Gold
Specialist[12] HEI	5	12	13
HEI high tariff	4	14	13
HEI medium tariff	3	19	9
HEI low tariff	11	16	4
HEI	1	6	4
FEC/AP	32	49	16

[11] Tariff grouping definition applicable to English HEIs only. Based on the 93 participating English HEIs, 31 providers in each tariff score grouping. Up until 2016 the groups are defined as high: over 390 tariff points, medium: 280 to 390 tariff points, low: under 280 tariff points.
[12] Providers which teach 80% of their activity within 3 cost centres.

From the model, estimates are obtained of the strength of association between the characteristics and the award, represented by the log odds ratio. Table 7 shows these for a selection of variables of interest, see Table A1 in Annex A for details of the effects of the other characteristics. The numbers inside the brackets are the standard errors of the estimates and represent the uncertainty associated with the log odds ratios. Large values indicate that greater uncertainty in the estimation of the strength of the association. A star denotes a statistically significant association (p-value less than 0.05).

Table 7 Estimates of the effects of a selection of characteristics on award from ordinal logistic regression model

Parameter	Coefficient (and standard error)	P-value	Odds ratio
Provider type = FEC/AP (HEI is the reference category)	0.104 (0.38)	0.78	1.11
Region of provider = London/South East (Elsewhere is the reference category)	-0.524 (0.28)	0.062	0.59
% students in most deprived category	1.325 (0.07)	< 0.0001	3.76*
% students who are not domiciled in the UK	0.729 (0.064)	< 0.0001	2.07*

What does this tell us?

For this model, an odds ratio above one indicates increased odds of getting a silver or gold award compared to bronze, with odds ratios below one indicate decreased odds of getting a silver or gold award compared to bronze. Odds ratios equal to one indicate that the odds of getting a silver or gold award is the same as the odds of getting a bronze award.

Using the model coefficients, one can consider what happens to the odds of getting a silver or gold award (compared to bronze) when a variable increases by one unit while all

other variables are kept constant. As an example, consider the coefficient for provider type where the coefficient is 0.104. As this value is on the log-scale, exponentiating 0.104 gives an odds ratio of 1.11, that is, there is an 11% (1.11 - 1 multiplied by 100) increase in the odds of getting a silver or gold award for FEC/APs compared to HEIs. However, in this case the p-value is 0.78 and so there is no evidence that this is a statistically significant increase in odds (as would be the case if $p<0.05$)

Now consider the region where the provider is located. Providers in London/South East region (compared to elsewhere in the country) are 41% (1 – 0.59 multiplied by 100) less likely to get a silver or gold award (compared to bronze). This coefficient is not significant ($p>0.05$).

Now consider the proportion of students from a deprived area. When the proportion of students from a deprived area increases by one unit, a provider is 276% (3.76 -1 multiplied by 100) more likely to get a silver or gold award (compared to bronze). This coefficient is significant ($p<0.05$).

Now consider the proportion of students not from the UK. When the proportion of students not from the UK increases by one unit, a provider is 107% (2.07 -1 multiplied by 100) more likely to get a silver or gold award (compared to bronze). This coefficient is significant ($p<0.05$).

Does provider type affect which award a provider (in London/South East) will get?

Table 8 shows the proportions of each award type for each provider type. It is noted that these are not calculated from the raw data but are the predicted proportions from the regression model and taken into account the possibly confounding of associations due to the relationships between the characteristics. In the results presented in Table 8, region is fixed as London/South East and the values of the other characteristics are kept at their mean. There is no discernible difference in the proportions of getting each award in FEC/APs and HEIs, and no statistically significant difference was detected within the model. Results from a model in which the small number of APs were excluded gave very similar results.

Table 8 Predicted proportions of each award type by provider type for providers in London/South East region. Values of the other characteristics are kept at their mean

Provider type	Bronze	Silver	Gold
HEI	0.34	0.53	0.13
FEC/AP	0.31	0.54	0.14

A more refined version of provider type was also considered, see Table 9 for the predicted proportions. It shows that the proportion of HEIs with low entry tariff getting a bronze award is higher than specialist HEIs or HEIs with high/medium entry tariff, however the p-value associated with this difference (p=0.23) indicates that the difference between the two proportions is not statistically significant.

Implication: The difference in awards between low tariff and high tariff providers have been taken into account in assessing evidence from the lessons learned exercise and in considering refinements to the process and metrics for next year and beyond.

Table 9 Predicted proportions of each award type by provider type and entry tariff for providers in London/South East region. Values of the other characteristics are kept at their mean

Provider type	Bronze	Silver	Gold
Specialist HEI	0.20	0.58	0.22
HEI high tariff	0.31	0.55	0.14
HEI medium tariff	0.32	0.55	0.14
HEI low tariff	0.49	0.44	0.07
HEI	0.27	0.57	0.16
FEC/AP	0.30	0.55	0.14

Are providers in the London and South East area more likely to get a bronze award?

Whether higher education providers in the London area are adversely affected by the TEF ratings is of particular interest as it is often cited (by providers) that there is a 'London effect'. Table 10 shows the predicted proportions of each award type by region. The results are presented for HEIs with the values of the characteristics kept at their mean. The proportion of providers located in London/South East region getting a bronze award is higher, and the proportion getting a gold award lower, when compared to other regions. However, this difference was not found to be statistically significant.

Implication: There is no evidence that any systematic regional issues have impacted the ratings. Therefore there will be no changes to the framework on regional issues.

Table 10 Predicted proportions of each award type for HEIs by region. Values of all other characteristics are kept at their mean

Location of provider	Bronze	Silver	Gold
Elsewhere	0.23	0.57	0.20
London/SE	0.34	0.53	0.13

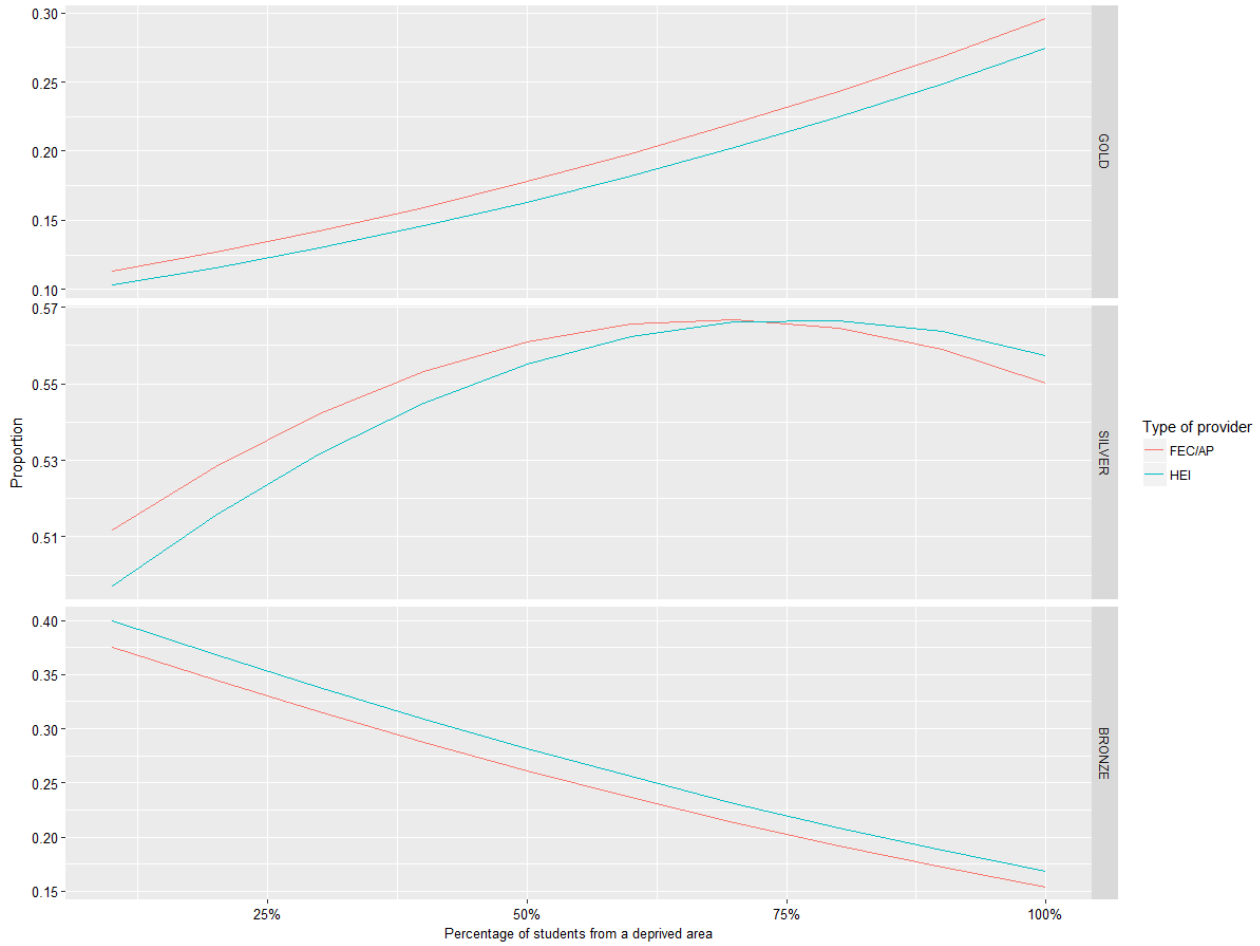
Does the proportion of students from a deprived area affect the type of award HEIs and FEC/APs can get?

Figure 2 shows, for each provider type, the relationship between the proportion of each award type and the proportion of students in the most deprived category. The results are presented for London/South East with all over characteristics kept constant (at their mean value). A decrease of 25% in predicted proportion of bronze awards is seen over the entire possible range of proportion of students who are in the most deprived category (from 0 to 100%), whereas for gold awards the opposite is seen. Within the range 0-75%, increases in the proportion of students who are in the most deprived category) are associated with an increase in the proportion of silver awards whereas for providers where the proportion of students from a deprived area is very high (>75%), the relationship changes with higher proportions of students from a deprived areas being associated with lower proportions of silver awards.

In summary, **having a high percentage of students from a deprived area does not increase a provider's probability of getting a bronze award, or decrease their probability of getting a gold award**. **This is true regardless of which region the provider is located.**

Implication: There is no evidence of an adverse affect of deprivation on the choice of final award. During the process that produces the benchmarks, a number of factors are taken into account, including deprivation. Therefore, it not necessary to make any further change to the framework based on this issue.

Figure 2 Predicted proportions of each award by percentage of students from a deprived area and provider type, for providers located in London/South East. Values for all other characteristics are kept at their mean



Does the proportion of providers getting each award differ for when considering the proportion of local students?

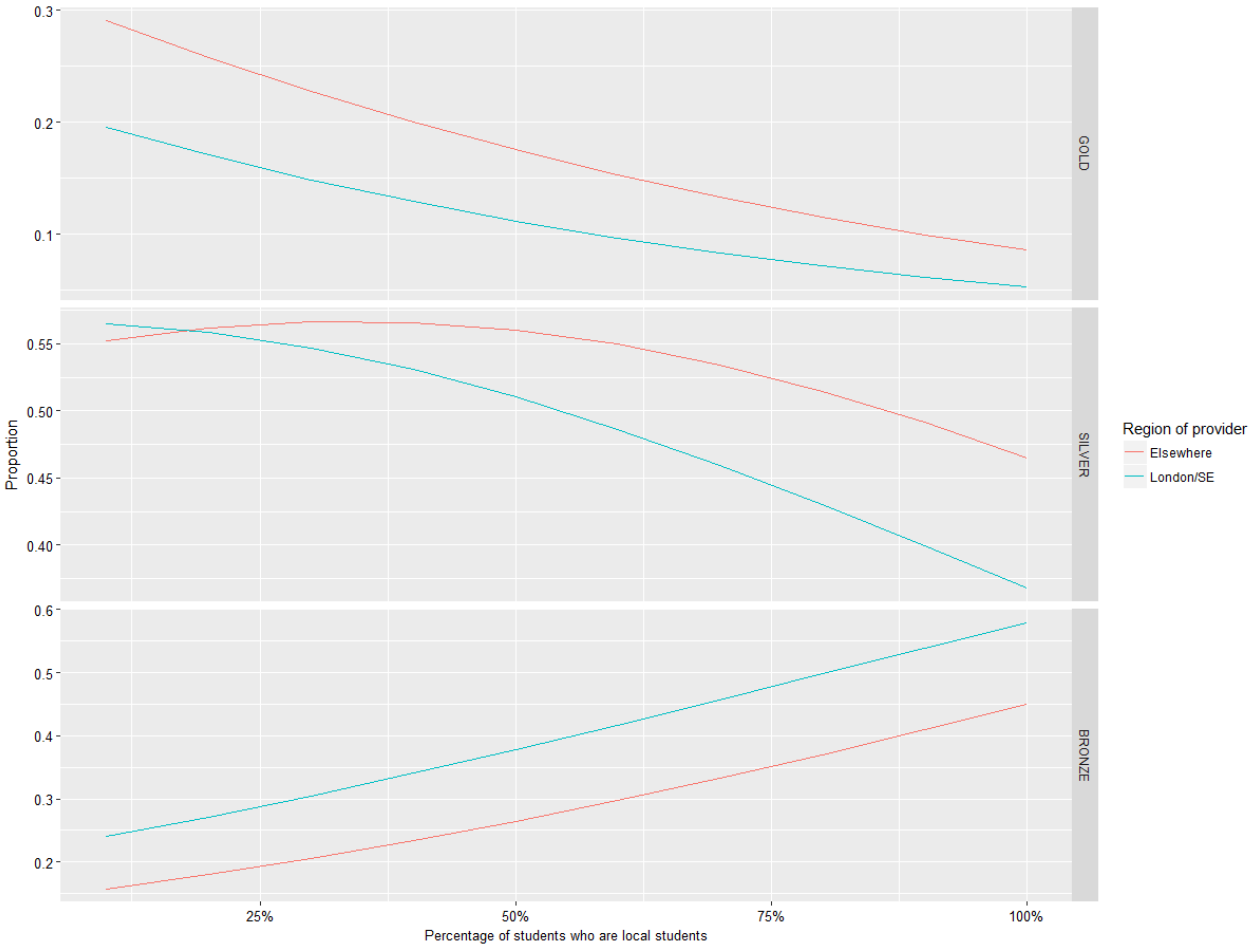
Figure 3 shows how the proportions of providers in each award type by the percentage of local students, for the region where provider is located. The type of provider is fixed as HEI. As the percentage of students who are local students increases, the proportion of bronze awards increases by between 0.3 and 0.35 for providers located in London/South East and elsewhere. The proportion of gold or silver awards decreases as the percentage of local students increases. When a provider has a low percentage of local students, the proportion of silver awards is slightly greater for providers in London/South

East area than providers located elsewhere. The decrease in the proportion of silver awards is steeper for providers located in the London and Southeast.

In summary, **having a high percentage of students who are local students increases a provider's probability of getting a bronze award and decreases their probability of getting a gold award regardless of which region the provider is located. This is confirmed by the model which shows that the percentage of students who are local students is statistically associated with the final TEF award (p-value < 0.05). The model also shows that region is not statistically associated with the final award (p-value > 0.05).**

Implication: The correlation between final TEF outcome and local students have been noted and will be taken into account when considering refinements to the process and metrics for next year and beyond.

Figure 3 Predicted proportions for HEIs by percentage of students who are local students and provider location. Values for all other characteristics are kept at their mean



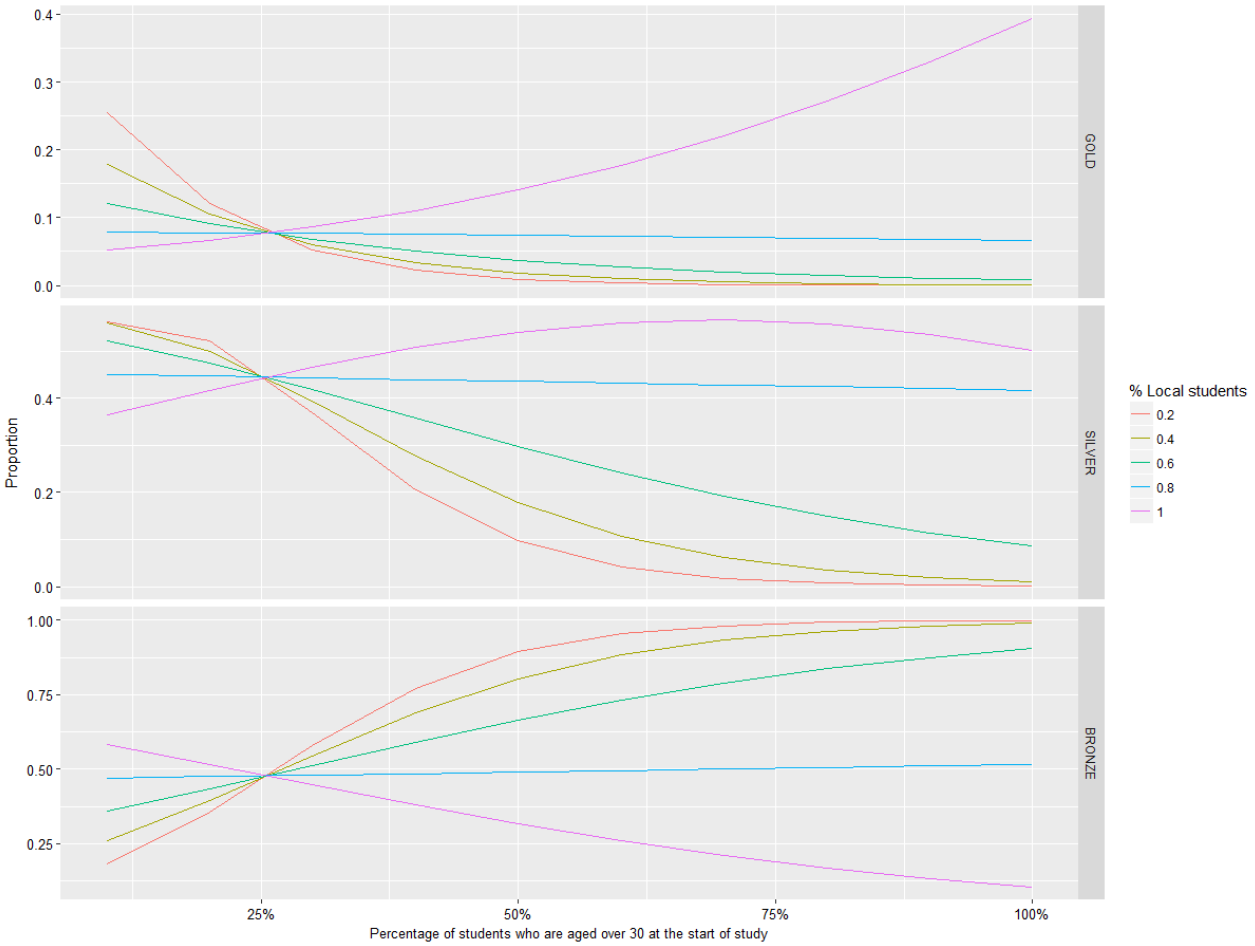
Is there a relationship between award and the levels of local students and older students?

Figure 4 shows how the proportions of each award type varies by the percentage of local students and by the percentage of students who are aged over 30 at the start of study. The type of provider is fixed as HEI. Region is fixed as elsewhere and all other continuous variables are kept at their mean. As the percentage of older students increases, the proportion getting a bronze award also increases for all except providers who have the highest percentage of local students where the proportion of bronze awards decreases. The proportion increases more sharply for providers who have a low percentage of local students. The proportion of silver awards decreases as the percentage of older students increases in particular for providers with a low percentage of local students. The proportion only slightly decreases for providers with the largest percentage of local students. The proportion of gold awards increase for providers with the highest percentage of local students as the percentage of older students increases. But for those providers with lower percentages of local students, the proportion of gold awards decreases as the percentage of older students increases.

In summary, **the proportion of providers getting a bronze award is higher in those with a high percentage of older students who are not local students than those with a high percentage of older students who are local . The model confirms that both of these characteristics are statistically associated with the final TEF award (p-values < 0.05).**

Implication: The correlations between final TEF outcome and older and local students have been noted and will be taken into account when considering refinements to the process and metrics for next year and beyond.

Figure 4 Predicted proportions for HEIs in London/South East area by percentage of students who are local students and percentage of students who are aged over 30 at the start of study. Values for all other characteristics are kept at their mean



Annex A Model development and assumptions

The model

An ordinal logistic regression model is considered because the categories for the final award are ordinal. There are three categories for award type (gold, silver and bronze) and for the model to work one category is compared against the other two. In an ordinal model, the probability of getting lower awards is compared to all awards above that.

In order to do this, two models are created:

- one that relates the characteristics to the log of the odds of getting a bronze award compared to silver or gold in a linear way;
- one that relates the characteristics to the log of the odds of getting a bronze or silver award compared to gold in a linear way.

The final models are given by,

$$\log\left\{\frac{P(award = bronze)}{P(award = silver\ or\ gold)}\right\} = constant_1 + b_1(provider\ type = FEC/AP) + b_2(region = London/South\ East) + b_3 Number\ of\ part\ time\ students + b_4\%\ students\ in\ most\ deprived\ category + b_5\%\ students\ who\ are\ aged\ over\ 30\ at\ the\ start\ of\ study + b_6\%\ students\ who\ are\ local\ students + b_7\%\ students\ who\ are\ not\ domiciled\ in\ the\ UK + b_8\%\ students\ who\ are\ not\ male + b_9(\%\ students\ who\ are\ aged\ over\ 30\ at\ the\ start\ of\ study \times \%\ students\ who\ are\ local\ students)$$

$$\log\left\{\frac{P(award = bronze\ or\ silver)}{P(award = gold)}\right\} = constant_2 + b_1(provider\ type = FEC/AP) + b_2(region = London/South\ East) + b_3 Number\ of\ part\ time\ students + b_4\%\ students\ in\ most\ deprived\ category + b_5\%\ students\ who\ are\ aged\ over\ 30\ at\ the\ start\ of\ study + b_6\%\ students\ who\ are\ local\ students + b_7\%\ students\ who\ are\ not\ domiciled\ in\ the\ UK + b_8\%\ students\ who\ are\ not\ male + b_9(\%\ students\ who\ are\ aged\ over\ 30\ at\ the\ start\ of\ study \times \%\ students\ who\ are\ local\ students)$$

Predicted proportions

Predicted proportions can be obtained by choosing fixed values for each provider characteristic included in the model and then using the parameter estimates to obtain the values:

$$\log\left\{\frac{P(award=bronze)}{P(award=silver\ or\ gold)}\right\} \text{ and } \log\left\{\frac{P(award=bronze\ or\ silver)}{P(award=gold)}\right\}.$$

By exponentiating these two values and re-arranging it is possible to get the probabilities for bronze and gold awards. Since the sum of the probabilities for the three categories sum to one, it is easy to derive the probability for a silver award.

Estimates of other model parameters

Table A1 shows the estimates of the model parameters. The numbers inside the brackets indicate how well the parameters are estimated. Large values indicate that the parameter has not been estimated very well. The star indicates that the parameter is statistically significant at the 5% level (p-value less than 0.05).

Table A1 Estimates of all model parameters

Parameter	Coefficient (and standard error)	P-value	Odds ratio
Constant: bronze/silver	-3.694 (0.24)	< 0.0001	0.025*
Constant: silver/gold	-1.124 (0.22)	< 0.0001	0.32*
Provider type = FEC/AP (HEI is the reference category)	0.104 (0.38)	0.78	1.11
Region of provider = London/South East (Elsewhere is the reference category)	-0.524 (0.28)	0.062	0.59
% students in most deprived category	1.325 (0.07)	< 0.0001	3.76*

Parameter	Coefficient (and standard error)	P-value	Odds ratio
% students who are aged over 30 at the start of study	-12.024 (0.081)	0.00	0.000006*
% students who are local students	-3.777 (0.61)	< 0.0001	0.023*
% students who are not domiciled in the UK	0.729 (0.064)	< 0.0001	2.07*
% students who are not male	-1.330 (0.15)	< 0.0001	0.26*
Number of part time students	0.00016 (0.00015)	0.28	1.0002

How the final model was selected

The deviance and Akaike Information Criterion (AIC) are used to compare competing models. The deviance is a quality of fit statistic that looks at the sum of the deviations between observed and predicted values. The deviance is usually used to compare a simpler model with a more complex model (for example, one with more independent variables). Under the null hypothesis that the simpler model is the true mode, then the difference between the deviances of the two models should follow a chi-square distribution.

The AIC is a model selection statistic that can be used to select the best model from a given set of models by looking at the loss of information in each model. As models can only approximate reality, the idea is to find a model that best approximates reality given the data that is available. AIC penalises extra unnecessary variables and therefore selects a model that fits well but is not overly complicated. The model with the smaller AIC is the better model.

The significance of each independent variable is also considered. The first stage of model development selects the one independent variable model with the smallest AIC and significant independent variable. The second stage adds a second independent variable to the already chosen independent variable (from the first stage) and selects the two variable model with the smallest AIC and significant independent variables. This

process continues until there is no further improvement in the model – either AIC does not change much or gets worse by the addition of further variables. These variables were considered during the variable selection stage but were discarded because they did not improve the model and were not significant: % students who are disabled; % students who are Black, Asian or other BME and a more refined version of region where provider is located.

The variables region (London/South East or elsewhere), provider type (HEI or FEC/AP) and the number of part time students are included in the model as baseline variables.

Model goodness of fit

Three goodness of fit tests are used to assess how good the model is at predicting the observed values. All three tests are used as each test detects different types of lack of fit.

The first is the Hosmer and Lemeshow test which splits the data in to 10 groups using predicted probabilities and creates observed and predicted frequencies for these 10 groups. From this a chi-squared statistic is created showing how the observed values differ from the predictions.

The second is the Lipsitz test which is similar to the Hosmer and Lemeshow test except, instead of using the predicted probabilities, each observation is given an ordinal score. If there are n observations in total (231 providers in our case), the observations are then organised in to g groups such that group 1 contains n/g observations with the lowest scores and group g contains the n/g observations with the highest scores. Indicator variables are then created with values equal to 1 if the observation is in that group and 0 otherwise. A new regression model is then fitted which includes these indicator variables. The log likelihood of this new model is then compared to the log likelihood of the original model.

The third test is the Pulkstenis and Robinson test. This test partitions data according to observed covariate patterns using the categorical independent variables only. Any unobserved covariate patterns are discarded. Each observation is assigned an ordinal score and the covariate patterns are split in to two groups based on the median ordinal score within each pattern. A table of observed and estimated frequencies are then constructed and a chi-squared test is used to assess the difference between observed and estimated frequencies.

The p-values from the three tests respectively are 0.48, 0.67 and 0.73. All are above 0.05 indicating that there is no evidence that the observed and expected frequencies differ.

Model assumptions

The main assumptions for an ordinal logistic regression model are:

1) The dependent variable is measured on an ordinal scale – this is true as final award (bronze, silver, gold) is clearly ordinal.
2) The independent variables are either continuous, categorical or ordinal – this is true as the model consists of a mixture of continuous and categorical contextual variables.
3) No multi-collinearity. Multi-collinearity can occur when two or more independent variables are highly correlated with each other. None of the independent variables are very highly correlated with each other (see Table 3). The largest correlation is 0.6 between proportion of students who are local students and proportion of students who are aged over 30 at the start of study. A formal test for detecting multi-collinearity looks at the condition indices with values greater than 30 indicating a problem. The condition index represents the collinearity of combinations of variables in the dataset. None of the condition indices for the model are above 30 so can conclude that multi-collinearity is not an issue for this dataset.
4) Proportional odds. In the proportional odds model, each outcome has its own intercept, but the same regression coefficients. This means the overall odds of any event can differ, but the effect of the independent variables on the odds of an event occurring in every subsequent category is the same for every category. To check if the proportional odds model is suitable we can compare it to the multinomial logistic regression model. This model does not take into account the ordering of the dependent variable and does not assume proportional odds. The model also has many more unknown parameters to estimate. If we assume that the proportional odds model is the better model, we can compare the difference of the deviance from each model with a chi-square distribution.

    Null hypothesis: proportional odds model is suitable

    Deviance for proportional odds model: 435.08

    Deviance for multinomial model: 425.95

    If the null hypothesis is true, then the difference (9.49) should follow a chi-square distribution with 9 degrees of freedom (the difference in the number of unknown parameters between the two models). The p-value from this test is 0.39 indicating that the proportional odds model is reasonable.

    In addition, the multinomial model leads to the same conclusions when examining how each provider characteristic affects the predicted probability of getting bronze, silver or gold. The multinomial model, however, is not as parsimonious as the proportional odds model as it contains many more unknown parameters and suffers from loss of information as it does not account for the ordering of the dependent variable.

Annex B Peer review

The analysis in this report has been peer reviewed by Professor Gavin Shaddick, Chair of Data Science and Statistics at the University of Exeter.

Is the technical methodology appropriate for achieving the objectives of the research? For instance, are the proposed statistical techniques and model specification adequate? Is the analysis robust?

The choice statistical methodology used in this analysis is appropriate and reflects the nature of the data and the aims of the analyses. The analysis has been performed systematically with careful examination of whether modelling assumptions are tenable and the possible sensitivity of the results to the specific choice of statistical model being assessed by using alternative methods where appropriate.

Are the variables appropriate for achieving the objectives of the research?

The overall objective of the analyses is to examine whether there (statistically significant) associations between award and provider type, region and student characteristics. There may be other factors that may be of interest, but for the current aims and based on the data that is available, then the variables used here, and the way in which they are used, are appropriate in the context of the aims of the analyses.

Do you agree with the interpretation of the results, as set out in the report? Are the conclusions too strong or need further testing/revising?

The findings from the initial (univariate) analyses and the interpretation of the coefficients from the (ordinal logistic) regression are accurately reported. In the executive summary the results and statistical significance (or not) are clearly stated with the latter used to provide clear guidance as to why apparent differences are not suitable for a basis for decision making.

Please include any other comments or suggestions, not covered above, here

Department for Education

Reference: DFE-RR735

ISBN: 978-1-78105-815-2

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications