

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/107497/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Gillard, Jonathan William 2017. An initial analysis and reflection of the metrics used in the Teaching Excellence Framework in the UK. *Perspectives: Policy and Practice in Higher Education* 22 (2) , pp. 49-57. 10.1080/13603108.2017.1409669

Publishers page: <http://dx.doi.org/10.1080/13603108.2017.1409669>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



ARTICLE TEMPLATE

An Initial Analysis and Reflection of the Metrics used in the Teaching Excellence Framework in the UK

J. W. Gillard

Cardiff School of Mathematics, Cardiff University, Senghennydd Road, CF24 4AG, UK

ARTICLE HISTORY

Compiled November 14, 2017

ABSTRACT

In this paper we look at the results from the recent Teaching Excellence Framework (2017), which were made publicly available in June 2017. We offer some initial analysis and commentary, look at the primary reasons for providers being awarded Bronze, Silver and Gold, and look at some providers close to the borderline for their award. We demonstrate that the Provider Submissions, a narrative document prepared to accompany the submission would have had a significant effect upon the award bestowed.

KEYWORDS

Teaching Excellence Framework; National Student Survey; metrics

1. Introduction

In this paper we offer some initial analysis and commentary on the metrics used in the Teaching Excellence Framework (TEF), the results of which were made publicly available on 22nd June 2017. TEF is a quality assurance exercise assessing the standard of education provision of higher and further education institutes in the UK.

The data investigated is available at HEFCE (2017), and is known as the so-called “TEF Year Two” data. The six metrics used for TEF and the data sources from which the metrics are computed are:

- Teaching on my course (National Student Survey)
- Assessment and feedback (National Student Survey)
- Academic support (National Student Survey)
- Non-continuation (Higher Education Statistics Authority and Individualised Learner Records)
- Employment or further study (Destination of Leavers from HE survey)
- Highly skilled-employment or further study (Destination of Leavers from HE survey)

Each metric is reported at so-called core or split level. At core level, each metric is considered for full-time and part-time students. The split level allows for more granularity of the core levels, splitting the core levels of full-time and part-time students into categories according to ethnicity, sex, disability and social disadvantage (as well

as others). The primary focus in this paper will be the core metrics for the majority mode of study for each provider, which is overwhelmingly full-time students (this is aligned to the assessment process of TEF and is described in more detail next).

The general procedure for TEF consists of three steps, and are described below. This text is largely verbatim from the specification document available at Department for Education (2017).

Step 1: Assessors review core and split metrics

- Assessors start by reviewing a provider's core metrics at provider level for the majority mode of study
- They also review the split metrics
- Assessors form an initial hypothesis of a rating based on performance against the metrics

Step 2: Assessors review the provider submission

- Assessors then look at the provider's submission
- They test the initial hypothesis to see if there is anything that causes them to take a different view of their initial rating

Step 3: Assessors review the provider's performance holistically

- Assessors then look holistically at their judgements - both performance against the criteria demonstrated by the metrics and the submission - using the descriptors
- They consider whether their judgement remains the same or should be adjusted accordingly

In this paper we primarily focus on the core metrics which underpin the assessment process. A provider entering the TEF may receive one of three possible awards: Gold, Silver or Bronze. Each provider is evaluated on the metrics given above, by comparing the values obtained against a benchmark value, primarily for the core metrics at the majority mode of study. A unique benchmark is calculated for each provider's core and split metrics. As stated in the specification document at Department for Education (2017), "the benchmark is a weighted sector average where weightings are based on the characteristics of the students at the provider". For each metric one flag from the set $\{--, -, =, +, ++\}$ is awarded. There are also some other flags awarded if there is insufficient data to evaluate a provider on a particular metric. Providers with three or more positive flags (either $+$ or $++$) and no negative flags are considered Gold. Providers with two or more negative flags (either $-$ or $--$) are considered Bronze and all other institutions are considered Silver. The final award of Gold, Silver or Bronze is formed at Step 3.

The flags $+/-$ are applied when the metric is at least $+/- 2$ percentage points from the benchmark and when the so-called Z-score is at least $+/- 2$. The flags $++/--$ will be applied when the metric is at least $+/- 3$ percentage points from the benchmark and when the so-called Z-score is at least $+/- 3$. The Z-score is defined to be the number of standard deviations the metric is from the benchmark.

We make the following notes of the analyses contained in this paper at the outset:

- (1) As outlined in Step 1 above, assessors start by reviewing a provider's core metrics

at provider level, focussing on the delivery mode in which providers teach the most students. This paper also follows this starting point, but will not comment on the split metrics directly.

- (2) We will be highlighting providers close to the borderlines for obtaining a different initial award of Gold, Silver and Bronze. This is as suggested by the data, and the measure of closeness to the borderline is open to discussion. This paper proceeds with a conservative choice.
- (3) This paper focuses on the publicly available data corresponding to the metrics detailed above. At the time of writing it difficult to assess the role or weighting assessors put on the accompanying providers submissions and metrics at the split level.

The structure of the paper is as follows. In Section 2 we give a high-level analysis of the TEF data, concentrating on the distribution of the Gold, Silver and Bronze awards amongst the providers. We also look at the distribution of positive and negative flags given, by metric. In Section 3 we look at features of providers awarded Gold, Silver and Bronze, and speculate on the main reasons why providers were given their awards. We also look at the vulnerability of providers (i.e. those close to the borderline of their given award).

2. Analysis of the data

2.1. Distribution of Gold, Silver and Bronze

There were a total of 233 entries to the TEF, consisting of three so-called provider types. The entries consisted of 134 Higher Education Institutions (HEIs), 93 Further Education Colleges (FECs) and 6 Alternative Providers (APs). Of these 233 entries, there were 55 Gold awards, 141 Silver awards and 78 Bronze awards given. Table 1 contains the percentage split of Gold, Silver and Bronze awards by provider type. A

	AP	FEC	HEI
Gold	33.33	15.05	32.09
Silver	50.00	51.61	50.00
Bronze	16.67	33.33	17.91

Table 1. Percentage split of Gold, Silver and Bronze awards by provider type

Pearson chi-square test on the counts yields ($\chi^2 = 12.049$, 4df, $p = 0.01699$) rejects the null hypothesis that there is independence between the provider type and the amount of Gold, Silver and Bronze awards given. A post-hoc analysis rejects the hypothesis that there is an independence between the Gold and Bronze awards given and the provider type ($p = 0.00717$).

2.2. Distribution of $\{-, -, =, +, ++\}$

Recall that each provider is assessed on six metrics, and for each metric a label from the set $\{-, -, =, +, ++\}$ is given. Figure 1 gives the percentage split of scores from $\{-, -, =, +, ++\}$ by provider type.

Figure 2 contains scatterplots of the percentage point deviation from the benchmark against Z-score, where each point (colour coded by award given) refers to a separate provider. The shaded rectangle in the bottom left hand corner refers to regions of the

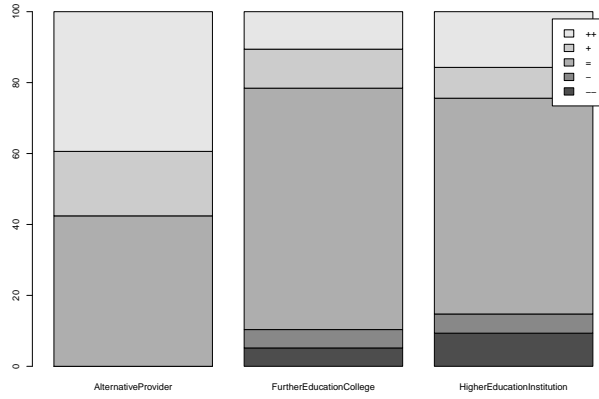


Figure 1. Percentage split cores on metrics by provider type

graph when a $-$ or $--$ will be awarded for the stated metric. The shaded rectangle in the upper right hand corner refers to regions of the graph when a $+$ or $++$ will be awarded for the stated metric. Naturally providers awarded Bronze are clustered at the bottom left hand corner of the scatterplots, and providers awarded Gold are clustered at the upper right corner of the scatterplots. Providers awarded Silver are more evenly spread across the range. There are some providers awarded Bronze that received high scores for Employment or Further Study, and similarly there are some providers awarded Gold who received low scores for this metric.

3. Vulnerability/features of providers obtaining a particular award

3.1. Primary reasons for providers obtaining Silver

To obtain a rating of Silver, providers need to receive only one negative flag (either $-$ or $--$). In this section we only consider those providers who were given an award of Silver, corresponding to the six metrics at the core level.

Table 2 contains the total numbers of scores awarded for each metric for all providers, and tables 3 and 4 split these results by provider type, namely Higher Education Institutions and Further Education Colleges. The amount of data for Alternative Providers is very small and so this split is omitted. The majority of negative flags are obtained from poor scores on the metric Highly Skilled Employment or Further Study, which interestingly is also the biggest provider of positive flags.

	$-$	$--$	$+$	$++$	$=$
AcademicSupport	5	7	9	6	87
AssessmentAndFeedback	8	10	9	20	67
EmploymentOrFurtherStudy	0	4	13	5	91
HighlySkilledEmploymentOrFurtherStudy	8	19	8	29	49
NonContinuation	6	7	8	9	85
TheTeachingOnMyCourse	9	4	8	3	90

Table 2. Distribution of $\{--, -, =, +, ++\}$ for providers awarded Silver

Table 3 shows that the primary reason for Higher Education Institutions being

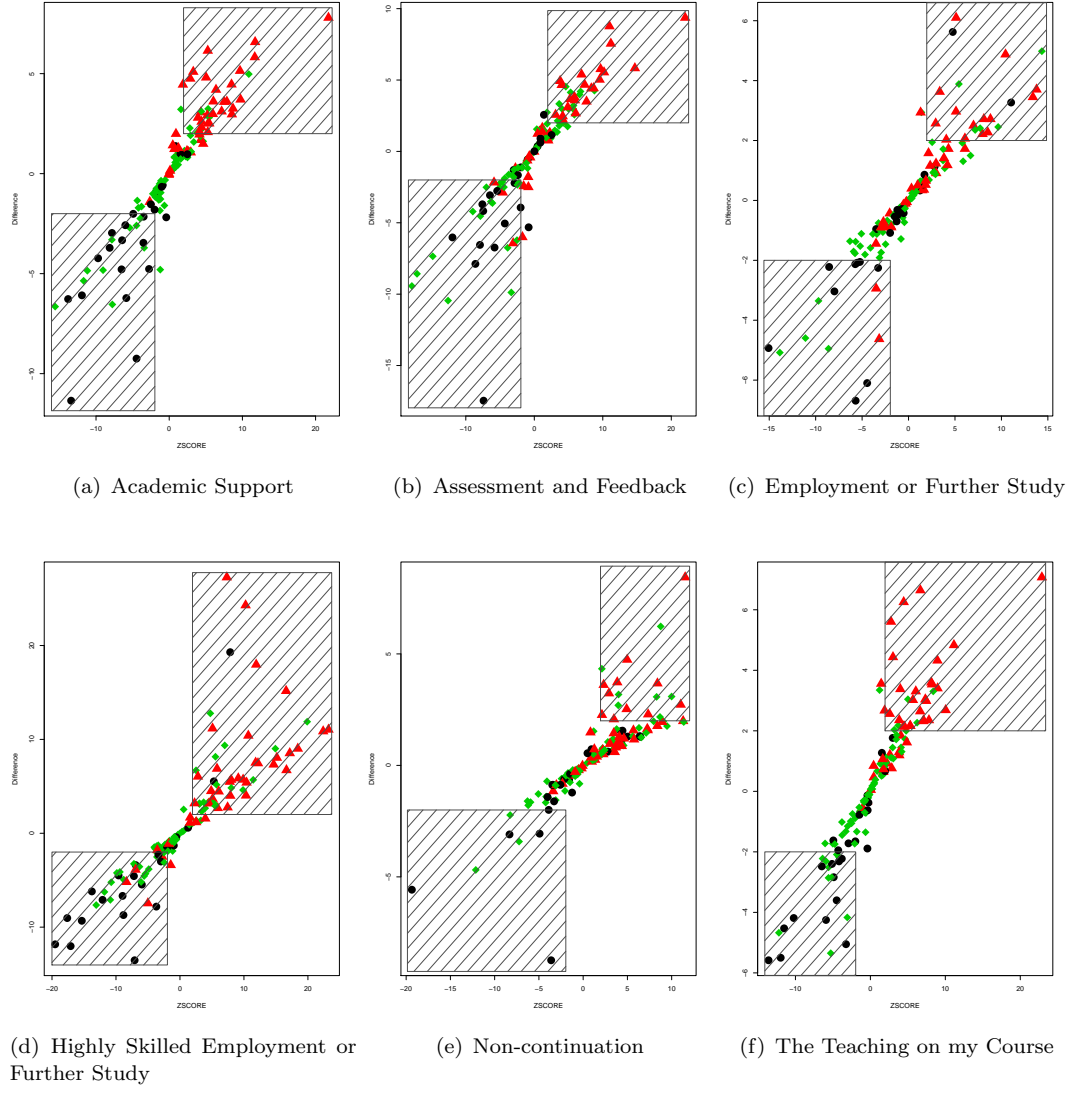


Figure 2. Scatterplots of the percentage point deviation from the benchmark against Z-score, by metric. Red triangles are providers that were awarded Gold, green diamonds are providers that were awarded Silver, and black circles are providers that were awarded Bronze.

awarded Silver is poor scores on Assessment and Feedback or Highly Skilled Employment or Further Study. Large amounts of positive flags were also awarded for Highly Skilled Employment or Further Study. Few negative flags are given for Non-continuation.

	—	--	+	++	=
AcademicSupport	3	7	4	3	50
AssessmentAndFeedback	8	10	5	7	37
EmploymentOrFurtherStudy	0	4	3	2	58
HighlySkilledEmploymentOrFurtherStudy	5	13	4	15	30
NonContinuation	1	2	3	4	57
TheTeachingOnMyCourse	7	3	4	2	51

Table 3. Distribution of {--, -, =, +, ++} for Higher Education Institutes awarded Silver

Table 4 shows that the primary reason for Further Education Colleges being awarded Silver is poor scores on Non-Continuation followed by Highly Skilled-Employment or Further Study. Unlike HEIs, more negative flags are given for Non-continuation. The reasons for HEIs and FECs being awarded Silver look to be different: the FECs awarded Silver did not receive negative flags for Assessment and Feedback, but this metric was a large source of negative flags for HEIs (for example).

	-	--	+	++	=
AcademicSupport	2	0	5	2	35
AssessmentAndFeedback	0	0	4	12	28
EmploymentOrFurtherStudy	0	0	10	3	32
HighlySkilledEmploymentOrFurtherStudy	3	6	4	14	18
NonContinuation	5	5	4	4	27
TheTeachingOnMyCourse	2	1	3	1	37

Table 4. Distribution of $\{-, -, =, +, ++\}$ for Further Education Colleges awarded Silver

Figure 3 contains scatterplots of the percentage point deviation from the benchmark against Z-score, where each cross refers to a separate provider. The shaded rectangle refers to regions of the graph when a $-$ or $--$ will be awarded for the stated metric. Recall that the flag $-$ will be applied when the metric is at least -2 percentage points from the benchmark and when the so-called Z-score is at least -1.96 . The flag $--$ will be applied when the metric is at least -3 percentage points from the benchmark and when the so-called Z-score is at least -3 . Given that the award of Silver will occur upon obtaining only one $-$ or $--$ flag, providers inside one shaded rectangle in the six graphs in Figure 3 will be given an Silver rating. The distance to the shaded rectangle, for each provider, could be viewed as a measure of the vulnerability of a provider on the stated metric. Such figures may be a useful management tool, to allow managers compare performance on a metric with other providers.

We can now look at the providers close to the shaded rectangles, which would suggest that these providers may have been close in obtaining a Bronze award. We define being close to the shaded rectangle as obtaining a percentage point deviation of between -1 and -2 and a Z-score between -1.645 and -1.96 . We note however that this is an arbitrary measure. The results are given in Table 5.

We now investigate providers who may have been close to achieving Gold. We consider the providers awarded Silver that received two positive flags ($+$ or $++$). It is important to acknowledge that there are likely to have been providers who may have been close to achieving Gold, such as those with one positive flag, and two metrics being close to receiving a positive flag (for example). These cases are less easy to distinguish from the data.

Figure 4 contains scatterplots of the percentage point deviation from the benchmark against Z-score where each cross refers to a provider. The shaded rectangle refers to regions of the graph where a $+$ or $++$ will be awarded for that metric. Recall that the flag $+$ will be applied when the metric is at least $+2$ percentage points from the benchmark and when the so-called Z-score is at least $+1.96$. The flag $++$ will be applied when the metric is at least $+3$ percentage points from the benchmark and when the so-called Z-score is at least $+3$.

We can now look at the providers close to the shaded rectangles, which would suggest that these providers may have been close in obtaining a Gold award. We define being close to the shaded rectangle as obtaining a percentage point deviation of between 1 and 2 and a Z-score between 1.645 and 1.95 . We note however that this is an arbitrary,

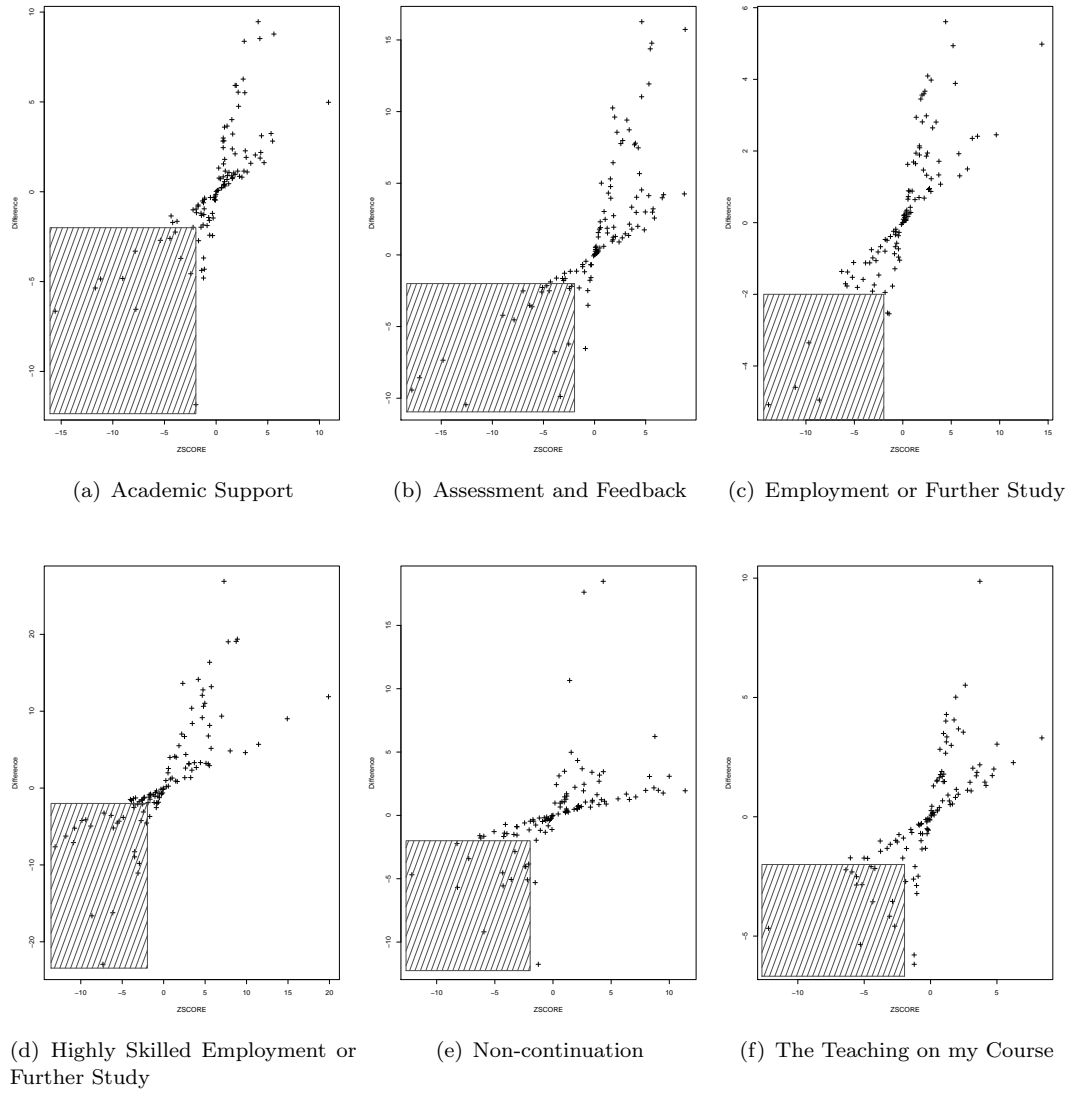


Figure 3. Scatterplots of the percentage point deviation from the benchmark against Z-score, by metric

yet conservative measure. The results are given in Table 6. There were no providers close to the shaded rectangle for the metrics not given.

3.2. *Primary reasons for providers obtaining Bronze*

In this section we only consider those providers who were given an award of Bronze, corresponding to the core data. Table 7 contains the total numbers of scores awarded for each metric for all providers, and tables 8 and 9 split these results by provider type, namely Higher Education Institutions and Further Education Colleges. The amount of data for Alternative Providers is very small and so this split is omitted.

Table 7 shows that the primary reason for being awarded a negative flag is poor scores on Highly Skilled Employment or Further Study.

Table 8 shows that the primary reason for Higher Education Institutions obtaining negative flags is poor scores on Highly Skilled-Employment or Further Study followed

Metric	Provider
Academic Support	Anglo-European College of Chiropractic, Cardiff Metropolitan University, Kingston College, Middlesex University, Oaklands College, Queen Mary University of London, St Mary's University (Twickenham), Strode College, Sunderland College, The Royal Agricultural University, University of Hull, University of St Mark and St John, Wrexham Glyndwr University, York College
Assessment and Feedback	University of Bradford, Redcar and Cleveland College, Ravensbourne, Oaklands College
Employment or Further Study	Anglo-European College of Chiropractic, Birmingham City University, Kirklees College, Redcar and Cleveland College, Strode College, University of Winchester
Highly Skilled Employment or Further Study	Hartlepool College of Further Education, Kaplan Open Learning (Essex) Limited, Kingston College, Leeds College of Art, North Lindsey College, University of Bolton, University of Abertay Dundee, University of Bristol, University of Worcester, Wrexham Glyndwr University
Non-continuation	Halesowen College, Hereford College of Arts, Leeds Beckett University, New College Stamford, Sunderland College, University of Central Lancashire, University of the Arts, London
The Teaching on my Course	Anglo-European College of Chiropractic, Furness College, Myerscough College, University of Bolton

Table 5. Providers close to the shaded rectangles in Figure 3, for each metric

Metric	Provider
Academic Support	Anglia Ruskin University, Swansea University
Employment or Further Study	Myerscough College, NCG, The Royal Agricultural University
Non-continuation	Bath Spa University, Oxford Brookes University, Sparsholt College, University of Hertfordshire
The Teaching on my Course	Stockton Riverside College, University of Sussex

Table 6. Providers close to the shaded rectangles in Figure 3, for each metric

	-	--	+	++	=
AcademicSupport	8	10	2	0	34
AssessmentAndFeedback	4	10	7	2	31
EmploymentOrFurtherStudy	8	6	4	3	33
HighlySkilledEmploymentOrFurtherStudy	5	21	4	4	20
NonContinuation	4	12	0	0	35
TheTeachingOnMyCourse	6	10	0	0	38

Table 7. Distribution of $\{-, -, =, +, ++\}$ for providers awarded Bronze

by Academic Support. There are very few positive flags awarded.

Table 9 shows that the primary reason for Further Education Colleges obtaining negative flags is poor scores on Non-continuation followed by Highly Skilled-Employment or Further Study. In comparison to Higher Education Institutions, more positive flags were awarded, but none were given for Non-continuation and the Teaching On My Course.

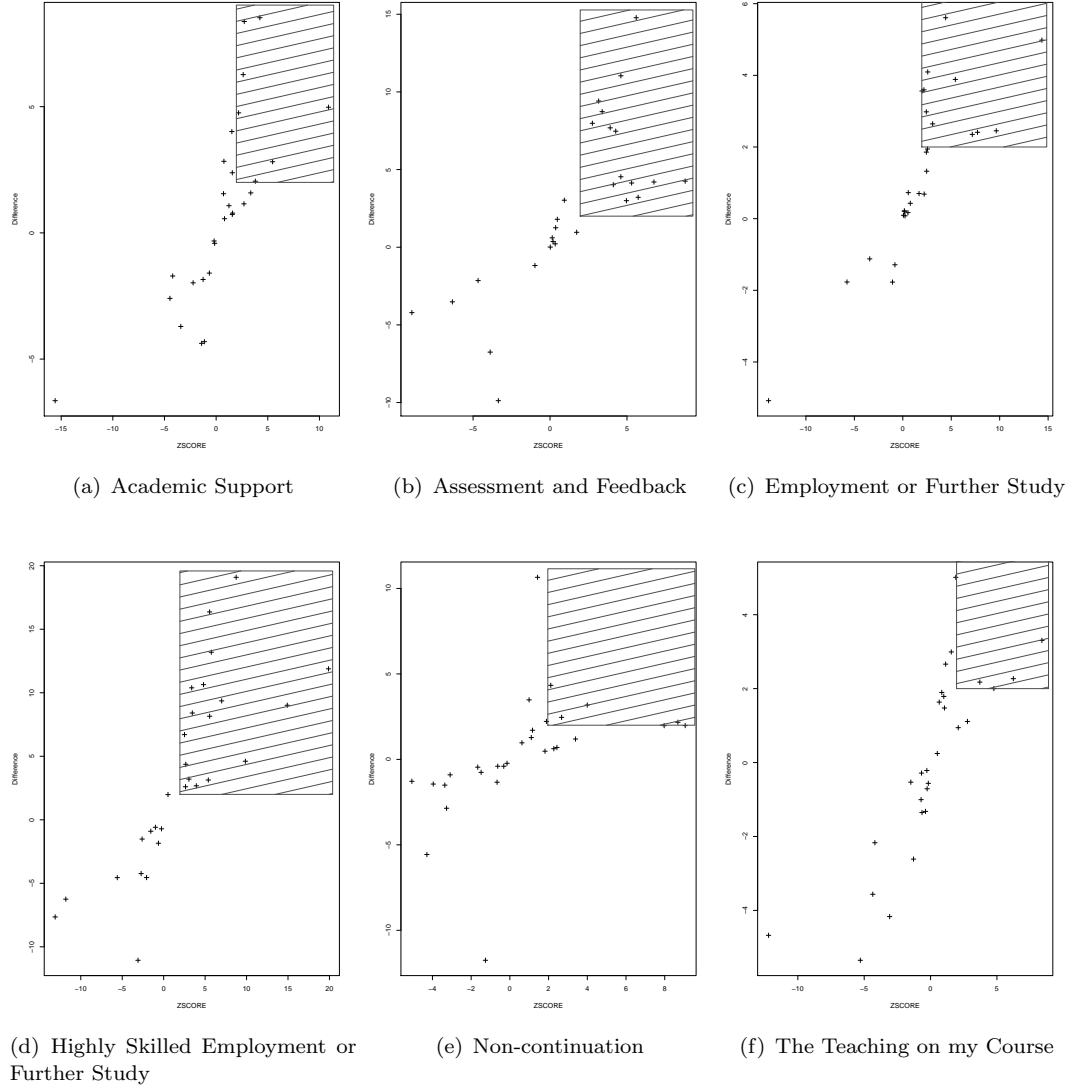


Figure 4. Scatterplots of the percentage point deviation from the benchmark against Z-score, by metric

	-	--	+	++	=
AcademicSupport	5	10	0	0	9
AssessmentAndFeedback	3	9	0	0	12
EmploymentOrFurtherStudy	5	4	0	2	13
HighlySkilledEmploymentOrFurtherStudy	2	14	0	2	6
NonContinuation	0	4	0	0	20
TheTeachingOnMyCourse	5	7	0	0	12

Table 8. Distribution of $\{-, -, =, +, ++\}$ for Higher Education Institutes awarded Bronze

Figure 5 is a histogram of the number of negative flags received by providers given a rating of Bronze, based on the core metrics evaluated on the majority mode of study. It can be seen that a considerable number of providers were awarded Bronze without a negative flag, or with one solitary negative flag (also see Figures 6(b) and (d)). This suggests that assessors did look beyond the core metrics to the metrics evaluated on

	-	--	+	++	=
AcademicSupport	3	0	2	0	25
AssessmentAndFeedback	1	1	7	2	19
EmploymentOrFurtherStudy	3	2	4	1	19
HighlySkilledEmploymentOrFurtherStudy	3	7	3	2	14
NonContinuation	4	8	0	0	14
TheTeachingOnMyCourse	1	3	0	0	26

Table 9. Distribution of $\{-, -, =, +, ++\}$ for Further Education Colleges awarded Bronze

the split data and/or reviewed the award as suggested by the data on the basis of the provider submissions. Alternatively these providers may have had insufficient data to be awarded anything other than Bronze, but on inspection this would only account for a small number.

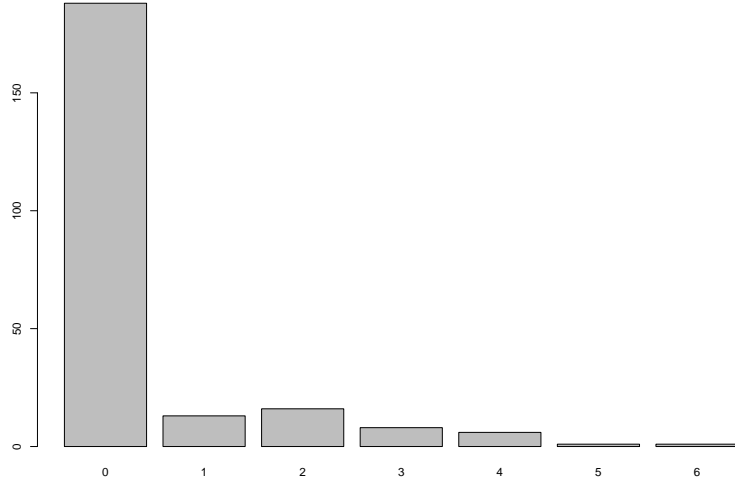


Figure 5. Histogram of the number of negative flags received by providers given a rating of Bronze

Figure 6 gives the histograms of the number of negative flags received by Higher Education Institutions and Further Education Colleges separately, with corresponding histograms omitting the providers awarded Bronze without any negative flags to clearer see the distribution. Ignoring those awarded Bronze with no negative flags, there are two Higher Education Institutions with five or six negative flags, but the mode number of negative flags for Higher Education Institutions awarded Bronze was two. There are two Further Education Colleges with four negative flags, but the mode number of negative flags for Further Education Colleges awarded Bronze was one.

3.3. Features of providers awarded Gold

Recall that providers with three or more positive flags (either + or ++) and no negative flags are considered Gold. In this section we only consider those providers who were given an award of Gold, corresponding to the core data.

Table 10 contains the total numbers of scores awarded for each metric for all providers, and tables 11 and 12 split these results by provider type, namely Higher

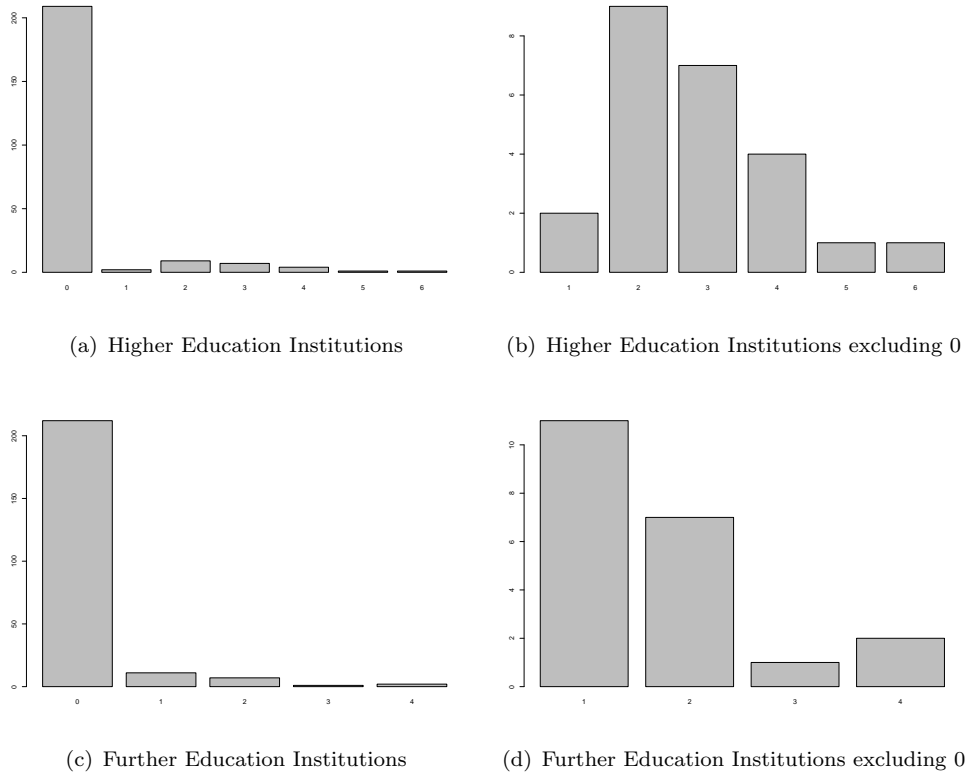


Figure 6. Histograms of the number of negative flags split by provider type, given a rating of Bronze

Education Institutions and Further Education Colleges. The amount of data for Alternative Providers is very small and so this split is omitted.

It is interesting to note that some providers received Gold, despite receiving negative flags, but the numbers of these are very small and are mainly Higher Education Institutes. We will investigate these Higher Education Institutes in more depth later in this section. Table 10 shows that providers awarded Gold rarely get negative flags, although there are some exceptions. Of all the metrics, scores on Non-continuation were the least likely to produce positive flags.

	-	--	+	++	=
AcademicSupport	0	0	14	25	20
AssessmentAndFeedback	2	1	8	30	18
EmploymentOrFurtherStudy	1	1	11	11	34
HighlySkilledEmploymentOrFurtherStudy	2	3	7	37	9
NonContinuation	0	1	11	6	41
TheTeachingOnMyCourse	0	0	15	17	27

Table 10. Distribution of $\{-, -, =, +, ++\}$ for providers awarded Gold

Table 11 gives the distribution of $\{-, -, =, +, ++\}$ for Higher Education Institutes awarded Gold. The highest amount of positive flags was given for Highly Skilled Employment or Further Study.

Table 12 gives the distribution of $\{-, -, =, +, ++\}$ for Further Education Colleges awarded Gold. The best source for positive flags was good scores on Assessment and Feedback, closely followed by Academic Support and Assessment and Feedback. Very

	-	--	+	++	=
AcademicSupport	0	0	11	16	16
AssessmentAndFeedback	2	1	5	18	17
EmploymentOrFurtherStudy	1	1	9	6	26
HighlySkilledEmploymentOrFurtherStudy	1	3	4	28	7
NonContinuation	0	0	8	5	30
TheTeachingOnMyCourse	0	0	13	13	17

Table 11. Distribution of $\{-, -, =, +, ++\}$ for Higher Education Institutes awarded Gold

few positive flags were given for Non-continuation.

	-	--	+	++	=
AcademicSupport	0	0	3	7	4
AssessmentAndFeedback	0	0	3	10	1
EmploymentOrFurtherStudy	0	0	2	4	7
HighlySkilledEmploymentOrFurtherStudy	1	0	2	8	2
NonContinuation	0	1	2	0	11
TheTeachingOnMyCourse	0	0	2	2	10

Table 12. Distribution of $\{-, -, =, +, ++\}$ for Further Education Colleges awarded Gold

Table 13 gives the distribution of $\{-, -, =, +, ++\}$ for Higher Education Institutes awarded Gold with at least one negative flag recorded on the core metrics. These results emphasize the importance of the accompanying written provider submission, as well as performance on the split metrics.

	-	--	+	++	=
Imperial College London	1	0	0	1	4
Liverpool Hope University	1	0	1	2	2
Rose Bruford College of Theatre and Performance	0	1	2	1	2
The Royal Central School of Speech and Drama	0	1	1	3	1
The Royal Veterinary College	1	1	1	0	3
The University of Northampton	0	1	2	1	2
The University of Nottingham	1	0	1	0	4
University of Derby	0	1	2	2	1

Table 13. Distribution of $\{-, -, =, +, ++\}$ for Higher Education Institutes awarded Gold with negative flags recorded on core metrics

4. Key messages for managers, administrators and policy makers

In summary, the main lessons that can be learned from the data are the following:

- There are significant variations in Further Education Colleges and Higher Education Institutes. Further Education Colleges suffer more with students ‘dropping-out but tend to receive higher scores for assessment and feedback, for example. Overall, in terms of negative flags awarded, it can be argued that Further Education Colleges performed better than Higher Education Institutes.
- The provider submissions had significant weight in improving the fortunes of a number of providers. For a significant proportion of submissions, the data and

guidelines suggest that the provider should have been given a Bronze or Silver award. On the basis of a good provider submission, then this was uplifted to Silver or Gold. Please note that there were few guidelines available to providers at the time describing what makes a good provider submission. The author understands that HEFCE are about to begin a detailed study of provider submissions.

- The primary reason for Higher Education Institutions being awarded Silver is poor scores on Assessment and Feedback or Highly Skilled Employment or Further Study. Large amounts of positive flags were also awarded for Highly Skilled Employment or Further Study. Few negative flags were given for Non-continuation. The primary reason for Further Education Colleges being awarded Silver is poor scores on Non-Continuation followed by Highly Skilled-Employment or Further Study.
- Higher Education Institutions awarded Bronze awards typically had poor scores on Highly Skilled-Employment or Further Study followed by Academic Support, whilst Further Education Colleges typically had poor scores on Non-continuation followed by Highly Skilled-Employment or Further Study.
- Naturally, providers awarded Gold generally performed well across all metrics, but reviewers did tolerate some poor scores in Highly Skilled Employment or Further Study. This evidences again the significant weighting that had to be given to the provider submissions.
- The methodology in this paper can be used to investigate the vulnerability, or excellence (relative to others) on each metric, highlighting good practise or areas needing improvement.

5. Conclusions

In this paper we have investigated the metrics used in TEF. We emphasize again, that the process for TEF is not purely data-driven: assessors allocate awards using the metrics as one component of their decision-making. The TEF process and data generated warrants further study, and this initial analysis has laid the foundations on which further analyses can be built.

One challenge in analysing TEF and its data is in the description of the assessment process (see Department for Education (2017)). It is stated that primary focus is initially given to the metrics on the core data and additional focus may be given to the split data, but precise weightings, and the explicit value of the provider submissions have not been given. Further clarification on the assessment process would be valuable.

References

- Department for Education. 2017. “Teaching Excellence Framework.” <https://www.gov.uk/government/collections/teaching-excellence-framework>.
 HEFCE. 2017. “TEF Year Two data.” <http://www.hefce.ac.uk/lt/tef/data/>.