TEF Year Three metrics questions and answers

Some of the cells in my workbook are blank or there is missing information. What's gone wrong?

When you open the Excel file from an email attachment, Excel may restrict editing capabilities. If this happens you will need to enable editing in the spreadsheet to bring back the formula-based functionality and see your results fully.

My numbers look much lower or higher than they should be – how have you attributed students to me?

The TEF Year Three metrics are based on a student's teaching institution, defined as the provider where a student spends the majority of their first year (or 2007-08, whichever is later). For franchised provision, students are included in the metrics of the teaching provider. The definition of the TEFUKPRN variable (available in the HEFCE TEF Year Three technical documentation at www.hefce.ac.uk/lt/tef/tefprocess/tech/) provides further detail.

For the contextual data, this will be shown as an average of the last three years, where available. Where only two years of data exist, the contextual data is averaged across these two years instead. Where only one year of data exists, this will be shown in the contextual data. Availability of data in any given year is determined at the overall cohort level, rather than being mode-specific. For example, if a provider has two years of part-time data, and three years of full-time data, all contextual data will be shown as the average of the last three years. This may mean that the annual averages for part-time appear lower than the provider may expect.

How have you identified my majority mode of provision?

In determining the number of years of suitable metrics that a provider has, HEFCE has identified the mode in which the majority of students are taught at the provider, as follows:

a. Only one majority mode has been calculated for each provider (rather than a majority mode calculated for each year of data). The majority mode has been calculated on the basis of the full time and part time student headcounts, averaged over the same number of years used for the provider's contextual data.

b. Where the headcount of full time students is greater than or equal to the headcount of part-time students, HEFCE has identified full time as the majority mode.

c. Where the condition above is not met and the headcount of part-time other undergraduate students is greater than or equal to the combined headcount of full-time and part-time first degree students, HEFCE has identified part time other undergraduate as the majority mode.

d. If neither condition b or c above is met the majority mode is part time.

A provider that has more than 35 per cent of students by headcount in its minority mode will be determined as having a similar number of student in both delivery modes. If this similarity exists for a provider, it will be clearly indicated on the metrics work book.

Why is my number of suitable years one rather than three?

The metrics are 'suitable' if each of the six metrics are reportable and benchmarked, either when aggregating all years of available data, or for at least one year.

If the metrics are not suitable on this basis, the 'number of years of suitable metrics' is zero.

If the metrics are suitable, the 'number of years' is then calculated individually for each of the six metrics as follows:

i. Where a metric **is** reportable and benchmarked **when aggregating all years of available data**, it is the number of years in which there are students **contributing** to that aggregated metric (this will be either one, two or three).

ii. Where a metric is **not** reportable and benchmarked **when aggregating all years of available data**, it is the number of individual years that are **reportable** and benchmarked (this will be either one or two).

The 'number of years of suitable metrics' is the lowest of these values across the six metrics for the majority mode.

The number of years of suitable metrics information is included in a provider's TEF Year Three metrics workbook.

How are the metrics defined? What contributes to the numerator and denominator for each metric?

Annex A of HEFCE's procedural guidance for Year Three provides a full description of each of the metrics. This includes detail on the categorisations that lead to a student being included in the numerator or denominator (or both) of each metric. The definitions of the variables available in the HEFCE TEF Year Three technical documentation (at www.hefce.ac.uk/lt/tef/tefprocess/tech/) provide further technical detail on these categorisations.

I have calculated the average of my indicator figures across the three individual years, and it doesn't match the core metric. Why not?

The core metric is simply the metric calculated using all three years' worth of data: it is an aggregate of the three years of data, rather than an average. For the indicator figures we total the three numerators and divide this by the total of the three denominators. Similarly, the benchmark is based on the aggregated cohorts.

How is the 'other undergraduate' level of study defined?

The TEF metrics include students who are included in the relevant Higher Education Statistics Agency (HESA) and Individualised Learner Record (ILR) datasets and registered on higher education (HE) Level 4, 5 and 6 programmes (or Level 6 only for the part-time continuation metrics).

Level 4 and 5 programmes are categorised as 'other undergraduate' in the descriptions of level of study within the TEF Year Three metrics. The categorisation is defined in the description of the TEFLEVEL variable in the HEFCE TEF Year Three technical documentation at www.hefce.ac.uk/lt/tef/tefprocess/tech/.

In broad terms, the 'other undergraduate' categorisation includes credit-bearing courses such as foundation degrees, diplomas and certificates of higher education, Higher National Diploma (HNDs)

and Higher National Certificate (HNCs), and undergraduate Professional Graduate Certificates of Education (PGCEs).

Why are the disadvantaged splits for the devolved administrations restricted to students from those administrations?

Appropriate measures of disadvantage have been set by each devolved administration. The measures chosen include the national indices of multiple deprivations (IMD), which are country-specific. In order to ensure consistent data, the populations need to be restricted in the same way; to do otherwise would risk performance being skewed by the different measures adopted in each nation. Where disadvantage is used in benchmarking the Participation of Local Areas (POLAR) measure is used consistently as it is the only UK-wide measure.

What is a benchmark? How do I use and interpret it?

Benchmarks are used to allow meaningful interpretations of provider-level statistics, by taking into account the different mix of students at each provider. In effect, we are considering the question: 'How well would the whole country do with this provider's students?' A unique benchmark is calculated for each provider's core, split and supplementary metrics.

The benchmark is a weighted sector average, where weightings are based on the characteristics of the students at the provider: it gives information about the sort of values that might be expected for that provider's indicator if the characteristics included in the weighting are the only ones that are important. The corollary of this is that where differences exist between a provider's indicator and benchmark, this may be due to the provider's performance, or to some other characteristic which is not included in the weighting.

The UK Performance Indicators and National Student Survey (NSS) outcomes already use this methodology.

The benchmarking process is a measure of indirect standardisation, or of us asking: 'What would the observed overall [continuation/satisfaction/employment] rate have been at this provider if its distribution of students across the unique benchmarking category combinations had been what it was, but its [continuation/satisfaction/employment] rates in those combinations were replaced by the national rates?'

How is a benchmark calculated?

A full explanation of the benchmarking methodology is given on the HESA website at https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks. A worked example of benchmarking is given at www.hefce.ac.uk/lt/tef/tefprocess/tech/.

A unique benchmark is calculated for each provider's core, split and supplementary metrics. This means that the provider is being compared with the whole sector, rather than with a pre-set group of providers. For the purpose of calculating benchmarks, 'the sector' is made up of all providers in scope for the TEF, regardless of whether they have met the eligibility criteria or chosen to apply for a TEF award.

The regional economy affects institutional employment outcomes. Why is region not included in the benchmarking?

While there are clearly regional effects on employment there are also many sub-regional effects which can be more pronounced. There are also technical reasons why including region in the benchmarks is problematic. The benchmarking for highly skilled employment or further study includes over 25,000 distinct groups, and adding region would increase this to over 300,000. This would significantly increase the extent to which providers' benchmarks were determined by their own students. The inclusion of context maps in the assessment is designed to allow assessors to make nuanced judgements based on local labour market conditions and the mobility of students at individual providers.

How are the entry qualification groups defined?

Where entry qualifications are included as a benchmarking factor, the specific groupings used are the categorisations defined in the descriptions of the TEFENTQUALGRP and TEFEMPENTQUAL variables, in the HEFCE TEF Year Three technical documentation at www.hefce.ac.uk/lt/tef/tefprocess/tech/.

The definition of the groupings used have sought to balance practical sense; homogeneity with respect to the indicator to which they refer; and appropriate spread of student numbers (including in conjunction with the other benchmarking factors used for the metric in question). As in the UK Performance Indicators, this means that the groupings differ between the metrics.

What is the standard deviation?

In general, small differences between an indicator and its benchmark are not important. However, it is not always obvious what constitutes a small difference. A standard deviation measures the amount by which one would expect a statistic to change, based solely on random sampling. It can therefore be used to say whether a difference between a provider's indicator and benchmark is significant or not. The standard deviations of the differences between the indicators and their benchmarks have been calculated using a method explained more fully on the HESA website, at https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks. They are the square root of the estimated variance of the differences.

Note that, because these are standard deviations of a statistic, they are more usually called standard errors.

The size of the standard deviation (of the difference between the indicator and the benchmark) for an institution is driven by the number of individual students across the institution. It is calculated by considering how different each individual student is from an 'average' student of the same type in the sector (based on their benchmarking factor characteristics). The standard deviation of the difference between a provider's indicator and its benchmark is then the square root of a function of the student-level differences from the 'average', adjusted for the number of students contributing to the relevant benchmarking groups.

In the standard deviation calculations described here, each of the student-level differences is therefore treated as a single observation. This means that the size of the standard deviation is unaffected by whether an institution has all of its students in a single benchmarking category, or each student in a different category. Broadly, as the number of students at an institution increases the size of its standard deviation will decrease.

What is the Z-score?

The Z-score is otherwise known as the standard score, the normal score or the standardised variable. 'Z' is used because the normal distribution is also known as the 'Z distribution'. It measures the number of standard deviations by which an observation is above the mean. The calculation of the Z-score tests whether the difference between the indicator and the benchmark is itself different from zero (meaning that the indicator and benchmark are not the same – rather, there is a statistical difference between them that does not arise by chance).

In general terms, the Z-score is obtained by subtracting the population mean from an individual score and then dividing this difference by the population standard deviation. This conversion process is called standardising or normalising. A positive standard score indicates above the mean, while a negative standard score indicates below the mean.

In TEF metrics terms, the Z-score is obtained by subtracting the institution's benchmark from its indicator, and then dividing this difference by the standard deviation defined above. The TEF flags will be applied such that:

- A difference of +2 percentage points **and** a Z-score of at least +1.96 will receive a positive flag, labelled '+'. If the benchmark is above 97 per cent the difference of 2 percentage points is not required.
- A difference of +3 percentage points **and** a Z-score of at least +3.00 the institution will receive a double positive flag, labelled '++'. If the benchmark is above 97 per cent the difference of 3 percentage points is not required.
- A difference of -2 percentage points **and** a Z-score below -1.96 will receive a negative flag, labelled '-'.
- A difference of -3 percentage points **and** a Z-score below -3.00 will receive a double negative flag, labelled '--'.

My Z-scores are impossibly large. How can this be right?

The Z-score tests the likelihood of the difference between an institution's benchmark and its indicator being due to chance. If the factors in the benchmarks accounted for all differences in institutional performance (such that all institutions were performing just like the sector and any observed differences were due to random variation alone) we would expect to see Z-scores fitting a normal distribution: an absolute Z-score of over 1.96 in 5 per cent of cases and over 3 in only 0.27 per cent of cases. However, if other factors not included in the benchmarks (such as institutional performance) are driving the differences we will see more atypical Z-scores because the causes of the differences are something other than chance alone.

More details of the effects of distributional assumptions on the Z-scores can be found in 'Statistical analysis of performance indicators in UK higher education' by D Draper and M Gittoes, in Journal of the Royal Statistical Society, Series A, volume 167, part 3, 2004: http://www3.unisi.it/eventi/dmq2006/paper/draper_gittoes.pdf.

As outlined in the description of standard deviations and Z-scores, the standard deviation for an institution's metric is likely to decrease as the number of students at the provider increases.

The use of standard deviations and Z-scores assumes an underlying normal distribution. Whether or not a student agrees, continues or gains employment is a binary outcome – how can assumption of a normal distribution be appropriate?

The answer to the previous question details our expectation that the Z-scores would fit a normal distribution if the factors in the benchmarks accounted for all differences in institutional performance (such that all institutions were performing just like the sector and any observed differences were due to random variation alone).

More details of the statistical model used to define the benchmarking approach and its system of flagging can be found in 'Statistical analysis of performance indicators in UK higher education' by D Draper and M Gittoes, in Journal of the Royal Statistical Society Series A, volume 167, part 3, 2004.

Null simulations undertaken by Draper and Gittoes – in which no underlying differences in 'quality' were present between institutions – identified that the distribution of Z-scores was close to standard normal, so that a flag based on |Z-score| > 3 was incorrectly raised only about 0.3 percent of time. This is what makes the use of the normal distribution appropriate.

What is the provider contribution to benchmark? How is it calculated?

The average proportion which the HE provider's own students contribute to the benchmark is provided as a context statistic. The statistic is designed to pick up situations where the benchmark is of limited use because few other HE providers are comparable. If the students at the HE provider contribute a large proportion to the benchmark, say more than 20 per cent, then the adjusted sector benchmark will be similar to the HE provider's own value.

Details of the calculation of this context statistic are available at https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks (under the heading 'Average contribution to benchmark').

Suppose that we are only including the factors of entry qualifications and subject of study in the benchmarking. The calculation of the context statistic is based on the sector grid of entry qualifications and subject of study. For the HE provider of interest, call the number of its students t, and let t_{ij} be the number studying subject i with entry qualification j.

To find the contribution of the HE provider's students to the benchmark, we use a weighted average of the proportion of each cell's students who come from the HE provider of interest. If the number of students in the sector who are studying subject i and have entry qualification j is T_{ij}, then in any cell the HE provider's students form a proportion $\frac{t_{ij}}{T_{ij}}$ of the total, and the context statistic is the weighted average of these values, namely:

$$\text{average contribution to the benchmark} = \sum_{\text{overall cells}} \frac{t_{ij}}{t} \times \frac{t_{ij}}{T_{ij}}$$

Is it possible to follow through the benchmark calculation?

HEFCE has published the sector average percentages for each of the metrics. This information is available at www.hefce.ac.uk/lt/tef/tefprocess/tech/. Providers should be aware, however, that the analysis required to recreate their benchmark figure will be substantial.

How will my benchmark change if I make an amendment?

The sector average percentages that have been used in the calculation of the benchmarks for each of the metrics have been fixed, and will remain as published at www.hefce.ac.uk/lt/tef/tefprocess/tech/. In the event of an amendment being approved, the amended profile of a provider's students with regards to the benchmarking factors will be compared with the fixed sector averages to recalculate their benchmark.

My metric has a high Z-score – why hasn't it been flagged with a '+' or '-' ?

Once the core and split metrics are calculated and benchmarked, those results that are **significantly and materially** different from benchmark are highlighted. If the difference between an indicator and its benchmark is small (less than 2 percentage points), the metric will not be flagged even if the Z-score is very high (more than 3). In such cases, the metric is only passing the (statistical) **'significantly different'** element of the test: it does not satisfy the '**materially different'** element.

Exceptionally, the materiality different test will not be applied. Where the benchmark is above 97 per cent and the provider's indicator is above the benchmark, the materiality test will not apply. In this circumstance core and split metrics will only have to meet the 'significantly different' test to be flagged.

My difference is very high, why hasn't it been flagged with a '+' or '-' ?

Once the core and split metrics are calculated and benchmarked, those results that are **significantly and materially** different from benchmark are highlighted. If the difference between an indicator and its benchmark is large (more than 3 percentage points), the metric will only be flagged if the Z-score is also high (more than 1.96). In such cases, the metric is only passing the **'materially different'** element of the test: it does not satisfy the (statistical) **'significantly different'** element.

My metric workbook shows a 'Yes' for 'Splits different to core metrics?', but the 'Core and metrics split' sheet shows that all the flags are the same (or all are blank). Surely the 'Yes' is wrong?

The analysis of whether splits are different from core metrics includes the detailed ethnicity sheet shown at the end of the workbook. In the scenario you have described, you should find that one of the detailed ethnicity splits is showing a different flag.

My metric workbook shows a Destinations of Leavers from Higher Education response rate of 75 per cent, and a denominator of 21. This would mean that 15.75 students responded to the survey which surely can't be right – 0.75 of a student can't have responded.

Firstly, note that the denominator is the denominator of the metric rather than the denominator of the response rate, which has a different definition, as explained below.

Students who respond to the Destinations of Leavers from Higher Education survey with an explicit refusal to answer any of the questions posed in the survey are counted as a response for the purposes of calculating the response rate. But these students **only** count towards the response rate calculation: because their response is 'explicit refusal' they do not contribute to the metric calculation. As per the definition of the TEFEMPINDPOP variable (available in the HEFCE TEF Year Three technical documentation at www.hefce.ac.uk/lt/tef/tefprocess/tech/), 'explicit refusal' responses to the DLHE do not meet the criteria to be included in the numerator or denominator for the metrics.

In the split metrics I have a mix of 'R', 'SUP', 'N/A' and other flags. Why do I have different flags across different splits?

The DLHE and NSS response rates shown in the metric workbooks are each:

a. The response rate to the whole survey (rather than to a specific scale of NSS questions)
b. The response rate observed in relation to the three-year aggregate core metric.

These response rates will help to determine whether the core NSS and DLHE metrics are suppressed or not.

However, in determining the split metrics the response rate (again, to the whole survey) of the population which informs each split is calculated and used as the basis for suppressing that split metric.

If the split metric does not trigger suppression on the basis of the survey response rate, we then consider whether or not any of the other suppression criteria have been triggered. If so, this suppression symbol will be shown (N if the indicator has a non-zero denominator of less than 10; N/A if the indicator has a denominator of zero; or SUP if the metric has insufficient benchmarking data). Otherwise, the metric flag will be shown ('++', '+', (blank), '-' or '--').

I have compared my TEF Year Three metrics with my results in the HESA UK Performance Indicators – why do they not match?

The TEF Year Three metrics are based on a student's teaching institution, whereas the HESA UK Performance Indicators are based on a student's registering institution. For franchised provision, students are included in the TEF Year Three metrics of the teaching provider whereas they would be reported in the registering institution's UK Performance Indicator (UKPI) result.

Depending on the metric in question, providers will want to take care in their interpretation of headline figures in the HESA publications. The HESA employment UKPIs are shown separately for first degree and other undergraduate qualifiers (Tables E1a and E1c). The HESA non-continuation UKPIs are shown separately for first degree and other undergraduate entrants (Tables T3a and T3d), each of which distinguishes between young, mature and all entrants.

In addition, users should be aware that the HESA non-continuation UKPIs are based solely on student records in the HESA student datasets. Students continuing in or transferring to active HE study recorded in ILR student records will be counted as an inactive outcome by the UKPIs, but as continuers in the TEF Year Three metrics. Similarly, students duplicated across HESA and ILR data will not be de-duplicated in the UKPIs but will be de-duplicated in TEF Year Three metrics.

Finally, we have not implemented the UKPI's institution-specific suppressions.

I have compared my TEF Year Three two metrics with my NSS results published by HEFCE and based on teaching institution – why do they not match?

The TEF Year Three metrics are based on a student's teaching institution, defined for the purposes of the TEF as the provider where a student spends the majority of their first year (or 2007-08, whichever is later). The definition of the TEFUKPRN variable (available in the HEFCE TEF Year Three technical documentation at www.hefce.ac.uk/lt/tef/tefprocess/tech/) provides further detail.

This differs from the definition of teaching institution used in annual publication of NSS results, which is based on the institution where the student is taught in their penultimate year of study.

We have reviewed our NSS target lists and found students who were not in their final year in one or more of the target lists being used in the TEF metrics. Can we now remove these students from that target list?

It is not possible to change the NSS target lists retrospectively. Such students were invited to participate in the NSS and may well have responded to the survey. The NSS process will now exclude these students from participating in future years of the NSS, so removing them from the target list and the TEF metrics would remove their voice.

I have compared our TEF Year Three metrics with our TEF Year Two metrics – why have they changed?

Potentially for a number of reasons: changes to the factors used in benchmarking and other changes introduced in the Department for Education's (DfE's) updated specification of TEF; changes to a provider or its data; and refinements to HEFCE's algorithms. The non-continuation metric has also been changed to measure continuation, to align its directionality with all other metrics (the underlying calculation has not changed).

Unless a provider has seen changes to its composition (via a merger) or its data, HEFCE expects that the indicator values of the Year1 and Year2 splits shown in the TEF Year Three metrics will differ only marginally, if at all, from their equivalents in TEF Year Two (the Year2 and Year3 splits respectively). Providers may see more substantive changes to the benchmark and associated Z-score values for these split metrics.

Changes to the factors used in benchmarking and other changes introduced within the DfE's updated specification of TEF

The TEF benchmarks have been amended slightly by DfE for the purposes of TEF Year Three. The principal changes that have been implemented are:

- Adding a new level of study benchmarking factor to all full-time metrics, and to all part-time employment-related metrics.
- Adding POLAR quintile, age and ethnicity to the benchmarking factors for continuation.
- Adding year as a new factor to the NSS-based metrics, reflecting the fact that some of the NSS questions changed this year.
- The subject of study benchmarking factor will no longer use subjects grouped by Joint Academic Coding System subject groups. The NSS-based metrics and the highly-skilled employment metrics (in both modes) use subjects grouped at Level 2 of HESA's Higher Education Classification of Subjects Common Aggregation Hierarchy. The employment and continuation metrics (in both modes) use nine groups of subjects, based on a slight

modification of the seven subject areas defined for the purposes of the TEF Year Three subject pilots.

- The entry qualification categorisations used in the benchmarking of continuation and employment-related metrics have been modified: students holding BTEC qualification have been disaggregated by grade within the benchmarking of continuation metrics, while the number of categorisations used in the employment-related metrics has been reduced to four.

In addition:

- The split metrics and context statistics for Welsh providers now show three categories instead of two, for the number of credits taught through the medium of Welsh.
- Providers in England and Wales now see two sets of split metrics for disadvantaged students: one based on POLAR and a second based on the IMD quintiles.

Changes to a provider or its data

If a provider has been involved in a merger since the release of the TEF Year Two metrics, this will be reflected in its TEF Year Three metrics. The HESA and ILR student data for all providers involved in a merger will be pooled together when calculating the TEF Year Three metrics, to ensure that the merged metrics reflect the makeup of the current legal entity.

In addition, if a provider has had HESA or ILR student data amendments approved by HEFCE specifically for the purposes of the TEF during Summer 2017, these have been included in the calculation of TEF Year Three metrics.

Refinements to HEFCE's algorithms

The TEF Year Three metrics are based on a student's teaching institution, defined for the purposes of the TEF Year Three as the provider where a student spent the majority of their first year or 2007-08, whichever was later. For TEF Year Two, a student's teaching institution was defined as the provider where a student spent the majority of their first year or 2009-10, whichever was later. This change accommodates the introduction of the supplementary LEO metrics, which consider the population of students who gained HE qualifications in academic year 2010-11.

The 'TEF Year Two questions and answers' document (www.hefce.ac.uk/lt/tef/tefprocess/tef2/) listed a small number of cases where HEFCE's TEF Year Two algorithms could be improved. These refinements have been made in constructing HEFCE's TEF Year Three algorithms.

This work has also sought to increase the alignment of algorithms as applied separately to HESA and ILR data to ensure consistency. We have combined the technical descriptions into one technical document for TEF Year Three, covering all sources of HESA and ILR student data. This means some fields have been renamed and some refinements have been made to the algorithms themselves. We believe that this work has strengthened our ability to link students accurately – to their DLHE and NSS survey responses in the cases of amended data returns, and to the most appropriate categorisation of their continuation outcomes when they traverse HESA and ILR data returns.

We have also extended and refined our use of linked ILR and National Pupil Database data to improve the identification of entry qualification grades and tariff points for those HE students recorded in ILR data.

Is TEFLOCPOSTCODE correct for records taken from the 2015-16 ILR?

The 'TEF Year Three metrics technical document' incorrectly stated in paragraph 159 that for the 2015-16 ILR return, TEFLOCPOSTCODE shows the value of HEPOSTCODE. However, this should instead state that TEFLOCPOSTCODE shows the value of HEPOSTCODE where it exists, and DELLOCPOSTCODE otherwise. The error was only in the technical document and not in the metrics themselves. The technical document was corrected on 8 November 2017.