

CONSULTATION RESPONSE

Teaching Excellence and Student Outcomes: subject level

May 2018

1. TO DEFINE 'SUBJECTS' IN SUBJECT-LEVEL TEF, DO YOU:

A) AGREE WITH USING LEVEL 2 OF THE COMMON AGGREGATION HIERARCHY AS THE CLASSIFICATION SYSTEM (CAH2, WITH 35 SUBJECTS), AND IF NOT, WHAT OTHER SYSTEMS COULD BE USED AND WHY?

B) THINK THAT SPECIFIC CHANGES OR TWEAKS NEED TO BE MADE TO THE DEFINITION OF THE 35 SUBJECTS IN CAH2, OR TO THE 7 SUBJECT GROUPS USED IN MODEL B, AND IF SO, PLEASE EXPLAIN WHY?

Using existing subject hierarchies and classifications is a more sensible option. This approach should mean that it will be easier to avoid duplication in data returns as higher education providers will already be collecting and collating that data for other purposes. As the CAH has been proposed by HESA following consultation and development, it would be unnecessary effort to go out to providers and other stakeholders again to establish different criteria. This would simply create excessive administrative burden, both in establishing these different criteria and in adapting systems to deal with them.

Nonetheless, it will be important to ensure that, should subject-level TEF develop according to model B, the 7 CAH groups have some flexibility to ensure that providers can manage instances of non-reportable data or a lack of data in small or new subjects.

2. DO YOU AGREE THAT WE SHOULD HAVE A LONGER DURATION AND RE-APPLICATION PERIOD IN SUBJECT-LEVEL TEF?

With both of the proposed models, submitting to the TEF will involve significant administrative effort from providers. The assessment exercise will also be extremely high-stakes, conferring reputational reward on providers and potentially (should legislation change) leading to differential income depending on ratings awarded. Compared to the REF, the rating awarded by the TEF has the potential to have a far greater impact on any provider, in terms of overall finance and recruitment.

Therefore, given that the REF is conducted once every 5-7 years, there is a strong argument for the TEF to take place over the same timescale. This is particularly true when considering the limited ability of universities to effect significant change over a short period of time, especially given the length of an undergraduate cohort is 3-5 years depending on the subject. A maximum duration between TEF assessments of 6 years is therefore a more reasonable proposition. Providers should have the right to re-submit during the first 3 years of the 6-year period.

3 SHOULD SUBJECT-LEVEL TEF RETAIN THE EXISTING KEY ELEMENTS OF THE PROVIDER-LEVEL FRAMEWORK (INCLUDING THE 10 TEF CRITERIA, THE SAME SUITE OF METRICS, BENCHMARKING, SUBMISSIONS, AN INDEPENDENT PANEL ASSESSMENT PROCESS AND THE RATING SYSTEM)?

The key elements of subject level TEF and provider level TEF should be as similar as possible.

4. FOR THE DESIGN OF SUBJECT-LEVEL TEF, SHOULD THE GOVERNMENT ADOPT:

- **A 'BY EXCEPTION' APPROACH (I.E. A FORM OF MODEL A), OR**
- **A 'BOTTOM UP' APPROACH (I.E. A FORM OF MODEL B), OR**
- **AN ALTERNATIVE APPROACH (PLEASE SPECIFY)?**

Without being able to consider the results of the subject-level TEF pilot, we cannot take a position on this matter at this stage. However, we can foresee at least one potential issue with each model that may prevent either from being taken forward without further work being undertaken.

Model A

The logic of model A is that a provider will receive, at the same time, an initial hypothesis at both provider level and subject level. It will also receive notification of which subjects are exceptions to its provider level assessment.

A provider will then submit a narrative to support or change the initial hypothesis. In the event of a narrative being used to improve the provider level submission succeeding, then it will change the subjects that are considered as by exception, thereby leading to wasted effort.

For example, a provider is awarded a silver overall in the initial hypothesis, with 6 of its subjects being considered as exceptions and awarded either bronze or gold. Should the provider believe that it has additional data and evidence to increase its provider level to gold, and submits to that effect while succeeding in that effort, the subjects originally not considered so - because the initial hypothesis was silver - become exceptions. However, the provider is not able to address these now exceptional subject areas in their submission.

Model B

Smaller providers, with small areas of provision or low numbers of students studying in subjects will face challenges with achieving reportable data. Providers who have recently established new areas of provision, such as clinical provision, will also be similarly challenged

In both of these cases, the provider 'runs out' of data. In the former example (smaller providers), even with 7 CAH groupings for subjects, it may be that the cohort is not large enough for the data to be considered robust, or it may be that the data is potentially identifiable (in regards to individuals).

In the latter example (newer areas of provision), it will take time to build up a longitudinal data profile that can be used to assess the quality of provision. This will be a particular issue where providers have responded to local need and are addressing regional skills issues. It will take at least 5 years to have enough NSS data to report, for example, and it would be at least 7 years before graduate outcomes data were available. This may, as an unintended consequence, disincentivise providers from broadening their subject portfolio significantly, lest they risk a reduction in their current rating or an expected increase in their award on their current subject portfolio.

5. UNDER MODEL A, DO YOU AGREE WITH THE PROPOSED APPROACH FOR IDENTIFYING SUBJECTS THAT WILL BE ASSESSED, WHICH WOULD CONSTITUTE:

A) THE INITIAL HYPOTHESIS RULE FOR GENERATING EXCEPTIONS FROM THE METRICS?

B) ALLOWING PROVIDERS TO SELECT A SMALL NUMBER OF ADDITIONAL SUBJECTS?

Both of these approaches are reasonable.

6. IN MODEL A, SHOULD THE SUBJECT RATINGS INFLUENCE THE PROVIDER RATING?

The key issue isn't necessarily that in model A the subject ratings have no influence on the provider, but rather that the exceptions are 'created' at the wrong point of the overall assessment, as outlined above.

However, if the logic of model A is to start with the provider rating and consider how subjects differ from it, then it is not sensible to then change that initial hypothesis solely on the basis of the subject ratings. Model B provides an opportunity for subject ratings to influence the provider rating. It seems odd to suggest that Model A replicates this, instead of what are now two very different approaches to create a subject-level TEF.

If the decision to allow subject ratings in Model A to influence the provider rating, then arguably this would also need to be piloted to understand how that process could work and what work it creates for the providers.

7. IN MODEL B, DO YOU AGREE WITH THE METHOD FOR HOW THE SUBJECT RATINGS INFORM THE PROVIDER-LEVEL RATING?

It is difficult to make a judgement about this method without further detail or examples. The one provided seems relatively straight forward – 4/7 of the subject groups have silver ratings, which logically suggests that the provider-level assessment would come out also as silver.

However, questions remain.

What if there are 2 subject groups with the same rating – e.g. 1 bronze, 3 silver, 3 gold – or ratings more evenly spread – e.g. 3/2/2? What impact does that have on the provider-level rating?

What if there are one or two very large subject groups that dominate the provider's offer in relation to undergraduate student numbers? Will their rating – because it is weighted by the number of students studying each subject – have a skewing effect?

8. DO YOU AGREE THAT GRADE INFLATION SHOULD ONLY APPLY IN THE PROVIDER-LEVEL METRICS?

MillionPlus would not agree with a policy of applying a metric of aggregate degree classification change (grade inflation) by which to assess institutions and so we disagree with its use in the TEF. Degree classifications are a matter for the provider to determine in line with their legal responsibilities as degree awarding institutions. The discharge of these awarding responsibilities are assured internally by the university and monitored externally on a co-regulatory basis by the Quality Assurance Agency with the Office for Students. This is part of the the 'Quality and Standards' strand of England's HE quality assurance landscape. Including a degree classification change metric in the TEF, which is a framework designed by the government ostensibly to encourage market choice by students, would not seem an appropriate setting for this measure. Degree classification issues should continue to be dealt with solely in the Quality and Standards strand of the (co)-regulatory framework.

9. WHAT ARE YOUR VIEWS ON HOW WE ARE APPROACHING POTENTIAL DIFFERENCES IN THE DISTRIBUTION OF SUBJECT RATINGS?

The proposal to allow a natural distribution within subjects is appropriate.

10. TO ADDRESS THE ISSUE OF NON-REPORTABLE METRICS:

A) DO YOU AGREE WITH THE PROPOSED APPROACH?

B) WHEN ASSESSMENT OCCURS, DO YOU PREFER THAT ASSESSORS:

- **RELY ON GROUP METRICS ALONGSIDE ANY REPORTABLE SUBJECT-LEVEL METRICS?**
- **RELY ON PROVIDER METRICS ALONGSIDE ANY REPORTABLE SUBJECT-LEVEL METRICS?**
- **FOLLOW AN ALTERNATIVE APPROACH (PLEASE SPECIFY)?**

Smaller providers, with small areas of provision or low numbers of students studying in subjects will face challenges with achieving reportable data. Providers who have recently established new areas of provision will also be similarly challenged.

In both of these cases, the provider 'runs out' of data. In the former example (smaller providers), even with 7 CAH groupings for subjects, it may be that the cohort is not large enough for the data to be considered robust, or it may be that the data is potentially identifiable.

In the latter example (newer areas of provision), it will take time to build up a longitudinal data profile that can be used to assess the quality of provision. This will be a particular issue where providers have responded to local need and are addressing regional skills issues. It will take at least 5 years to have enough NSS data to report, for example, and it would be at least 7 years before graduate outcomes data was available. This would apply to any of the new medical schools that have been recently approved, or any provider that has moved into areas of STEM provision or digital skills to address regional skills gaps. Even if they are meeting the need appropriately and to the satisfaction of employers and students, the TEF will not be able to evidence this.

The issue of non-reportable metrics therefore becomes a sequential one. Providers and subjects will not be fully fit for assessment for many years, which is likely to put them at a disadvantage compared to providers with long established provision. This is in contradiction with the government's policy objective to create more choice and competition for students. Again, this may discourage providers from changing significantly their subject portfolio offer.

There are at least two potential ways to address non-reportable metrics in addition to the ones proposed. Firstly, if Model B is selected, it may be possible to allow providers to define subject groups that are large enough to return reportable metrics. In many cases, the 7 groupings that currently exist may be sufficient, but an option to create larger groups should be available to providers. There is already criteria for reportability; providers should be given some discretion to create groups that are large enough to create reportable data.

A second option is that providers or subjects are awarded provisional ratings based on non-reportable metrics, with appropriate caveats noted alongside the rating. This would be beneficial in respect of new areas of provision, enabling the provider to have a rating while it builds a longitudinal cohort that is fully reportable.

11. Do you:

A) AGREE THAT QAA SUBJECT BENCHMARK STATEMENTS AND PSRB ACCREDITATION OR RECOGNITION SHOULD REMAIN AS A VOLUNTARY DECLARATION, AND IF NOT, WHY?

It seems appropriate to maintain this as a voluntary declaration and allow providers to make the judgement of whether the information available is suitable for inclusion in a TEF submission. To introduce an external element diminishes the principle that providers are making the case for their provision.

B) THINK THAT THERE ARE ANY SUBJECTS WHERE MANDATORY DECLARATION SHOULD APPLY?

No.

12. DO YOU AGREE WITH OUR APPROACH TO CAPTURING INTERDISCIPLINARY PROVISION (IN PARTICULAR, JOINT AND MULTI-SUBJECT COMBINED COURSES)?

MillionPlus believes further work needs to be conducted before expressing a view. The proposed approach seems complicated, and will require students to deduce the value of a joint programme by looking at two separate ratings, which may be in entirely different subject groups, with entirely different approaches. It does not seem in keeping with the principle behind the TEF of providing potential students with meaningful information to support choice making with data that may be opaque and complex to understand.

13. ON BALANCE, ARE YOU IN FAVOUR OF INTRODUCING A MEASURE OF TEACHING INTENSITY IN THE TEF, AND WHAT MIGHT BE THE POSITIVE IMPACTS OR UNINTENDED CONSEQUENCES OF IMPLEMENTING A MEASURE OF TEACHING INTENSITY?

MillionPlus does not support a measure of teaching intensity in the TEF. Aside from in-principle objections about the approach - and concerns that all measures proposed so far are arbitrary and misleading - it is also a measure based on inputs. Therefore it is directly contradictory to the aims of the TEF – which includes the descriptor ‘Student Outcomes Framework’ in its full name – and the underpinning framework of the Office for Students, which is also focused on outcomes. If this is the new regulatory approach, then it is entirely inappropriate for an external assessment of teaching intensity that will be based on untested and uninformed measures. MillionPlus notes that the standard evidence report on quality in higher education, Graham Gibbs’s *Dimensions of Quality* (Higher Education Academy, 2010) does *not* list teaching intensity / contact time as one of its six key variables affecting teaching quality.

14. WHAT FORMS OF CONTACT AND LEARNING (E.G. LECTURES, SEMINARS, WORK BASED LEARNING) SHOULD AND SHOULD NOT BE INCLUDED IN A MEASURE OF TEACHING INTENSITY?

As stated above, MillionPlus does not see any basis for including a measure of teaching intensity in the TEF. Any measure that in some way suggests it is all about a certain type duration of contact would not be underpinned by evidence. It may be possible to measure intensity in some way, but for what purpose? Given all the variables involved in an individual’s learning experience, suggesting a link between the size of a class, the number of hours, or the seniority of a teacher (to take 3 of the given options) with attainment AND destination is highly problematic. We also fail to understand the arguments excluding any consideration of the general level of student independent learning, where there is an established correlation this and the student’s perception of the benefit they have gained from the programme of study. Independent learning is a fundamental element of the university experience. Notwithstanding our objections to including a measure of teaching intensity in the TEF, suggesting independent learning should be ignored completely misunderstands and misrepresents the nature of university learning.

To take two examples.

1. In an intensive tutorial-based environment that requires high volumes of independent reading in order to submit regular assignments, the tutor is only briefly involved, yet highly influential to the learning that takes place.

2. In an environment that promotes problem-based learning, or encourages undergraduate research as part of the curriculum, the onus will be on individuals or groups of students to manage the workload under the supervision of an academic. That individual may not do what some consider 'teaching' but they will be influential in directing the learning that takes place.

A university hosts the intellectual space within which individuals can acquire knowledge, develop their understanding and apply evidence and thought to problems and questions. Suggesting that this can be distilled to a measure of contact hours or class size fundamentally damages the notion of independent enquiry and might discourage legitimate innovations with blended learning, some of which might involve a reduction in face-to-face contact time but no diminution of educational engagement between the provider and the student.

15 WHAT METHOD(S)/OPTION(S) DO YOU THINK ARE BEST TO MEASURE TEACHING INTENSITY?

PLEASE STATE IF THERE ARE ANY OPTIONS THAT YOU STRONGLY OPPOSE AND SUGGEST ANY ALTERNATIVE OPTIONS.

N/A

16 DO YOU HAVE ANY OTHER COMMENTS ON THE DESIGN OF SUBJECT-LEVEL TEF THAT ARE NOT CAPTURED IN YOUR RESPONSE TO THE PRECEDING QUESTIONS IN THIS CONSULTATION?

The consultation does not offer an opportunity to comment on the inclusion of longitudinal earnings data (LEO) in the TEF. There has been much speculation regarding this, particularly from government ministers and some commentators. However, the use of LEO data is highly problematic and risks significant unintended consequences. The use of LEO promotes the notion of a single or overriding value for a university education – that of the salary someone achieves within the first few years of graduation. Long-term impact, entrepreneurial activity that increases employment (but does involve a high income for the graduate), social value, or the importance of a skilled public sector workforce do not appear to be present in this earnings-focused approach. It is also misguided to assume that high salaries after graduation are a direct consequence of the course an individual studied, or the university they attended. The Institute of Fiscal Studies¹ finds that family background remains the major determinant of higher graduate earnings and career prospects even ten years after graduation.

Including LEO data in TEF as an assessment of the quality of teaching at university could undermine key sectors of the economy by creating disincentives for courses that are considered "less value for money" according to an earnings metric. This could adversely affect any attempts to address the productivity and skills challenges facing the country harm the creation of a robust industrial strategy and a rebalanced economy. Moreover, students from more disadvantaged backgrounds, who are shown to be more debt-averse than their more affluent peers, could be incentivised to study certain subjects over others, because of earnings data, which would also impact negatively on social mobility.

The consideration of LEO data without proper context, as is currently the case, takes no account of the fact that earnings are lower in some sectors of the economy and that patterns of graduate employment differ. For example, graduates engaged in the UK's world-leading creative industries often start by developing portfolio careers and / or engaging with micro or small businesses where earnings are initially lower. Earnings of graduates working in the not-for-profit and in public services are unlikely to match those that can be gained

¹<https://www.ifs.org.uk/uploads/publications/wps/wp201606.pdf>

in other sectors, while earnings associated with courses attracting a majority of female students may be lower because of the gender pay gap.

A recent IFS study acknowledged the impact of context and geography on graduate salaries and concluded that 'Given regional differences in average wages, some very locally focused institutions may struggle to produce graduates whose wages outpace English-wide earnings, which includes those living in London where full time earnings for males are around 50% higher than in some other regions, such as Northern Ireland'.²

² <https://www.ifs.org.uk/uploads/publications/wps/wp201606.pdf>