Teaching Excellence and Student Outcomes Framework: Subject-level

A submission from the British Academy
May 2018

For further information, contact:
Maxime Delattre
m.delattre@britac.ac.uk
+44 (0)207 969 5318

Introduction

The British Academy is the UK's national academy for the humanities and social sciences. A Fellowship of over 1000 of the country's leading academics, the Academy received its Royal Charter in 1902. It exists to promote and champion its disciplines, and awards funding to researchers at all career levels. The humanities and social sciences provide a critical lens through which Government and society can address the wide-ranging challenges we face today.

The British Academy welcomes the opportunity to respond to this consultation.

The response has been produced with reference to available evidence and previous Academy positions, notably the Academy's response to the TEF technical consultation in May 2016.

The British Academy has asked Professor John MacInnes, in his role as Strategic Adviser for Quantitative Skills, to provide a more detailed commentary on the consultation document, which we are submitting as an annex to our response. This commentary addresses specifically questions 1 and 3 of the consultation document.

The Academy is primarily concerned that the metrics that the new framework intends to use to capture excellent teaching at subject and provider level are fundamentally flawed, and that attempts to measure either teaching intensity or grade inflation are highly problematic and may foster perverse behaviour. The current proposals also do not provide an adequate definition of interdisciplinarity or a robust process for evaluating interdisciplinary teaching excellence.

Q1. 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?

To meet the objectives of TEF in terms of providing accurate and insightful information on teaching quality to prospective students, the Government needs to select a classification system which strikes the right balance between a sufficient level of aggregation that would bear statistical significance, and a sufficient level of granularity.

None of the classification systems that currently exist are perfect. Because of the diversity of the departmental and teaching provision structures across providers, it will be impossible for any classification system to be aligned with the units and module structures within universities.

While the Common Aggregation Hierarchy (CAH2, with 35 subjects) appears a reasonable and acceptable compromise between the necessary levels of aggregation and granularity, it will be important for the panels to look very closely at, and take account of, the providers' written submissions to understand the context of the teaching provision for some courses, including interdisciplinary ones, where there is significant variance between courses within

the 35 subject areas. For instance, one would expect considerable difference in teaching within the broad category of "creative arts & design". Equally, one must consider how some disciplines may end up being split between different categories; for instance, in the case of archaeology, some BSc courses will be classified as "physical, material & forensic sciences" while BA courses fit within "history & archaeology" even though they are being taught in the same department. This may not be easily understood by prospective students when reading subject-level results.

The Academy is concerned that, should the process fail to allow panels to fully consider the institutional and departmental context of teaching in assessing and rating a subject, there is a significant risk that higher education providers will attempt to align their governance structures and aggregation of subjects to correspond to the new framework classification. This would constitute a costly reorganisation and a threat to innovation and diversity strategies within institutions. It could also lead to the closure of some study programmes or unhelpful pressure on some departments to conform to wider school or faculty policies which may not fit. Such standardization could harm the choice and diversity of provision available to students.

Q2. Do you agree that we should have a longer duration and re-application period in subject-level TEF?

The British Academy supports the Government's decision to extend both the award duration and the re-application period.

The Academy is concerned that a truly robust subject-level TEF, one that could gain the trust of the academic community, could soon grow to the scale and cost equivalent of Research Excellence Framework. The extension of the award duration and re-application period will contribute to containing these administrative costs and will reduce the burden on university staff.

In addition, it will also reduce the pressure on higher education providers to be constantly driven by TEF requirements and will allow them time to develop more innovative methods of teaching and governance structures for their courses. New courses take time to develop, and this process will have an impact on their results in metrics while providers respond internally to feedback.

Q3. Should subject-level TEF retain the existing key elements of the provider-level framework (including the TEF criteria, the same suite of metrics, benchmarking, submissions, an independent panel assessment process and the rating system)?

The Academy is concerned that the consultation document for subject-level TEF proposes to retain metrics from the provider-level framework that fail to provide an accurate picture of teaching quality. In its response to the Government's consultation on the Green Paper *Higher education: teaching excellence, social mobility and student choice* (January 2016) and on the *Teaching Excellence Framework: technical consultation for year 2* (July 2016), the Academy voiced its concerns regarding the use of these core metrics based on the NSS questions, stating that they are not fit for purpose, for both substantive and technical reasons.

The Academy welcomes the reduction in the weighting of the NSS metrics in TEF but remains skeptical of the value or reliability of the survey for assessing teaching quality. The results of NSS questions are based on student satisfaction, which often do not provide relevant or reliable information to measure the quality of programmes, as it represents nothing more than a snapshot of student feedback at a single point at the end of a degree. In addition, the metrics as they stand do not differentiate among the majority of providers, and there is a significant risk that they will not be able to differentiate among subjects and/or that such differentiation will result from omitted variables that are not measured or controlled for in the NSS.

Using multi-level analysis, Marsh and Cheng[1] show that only a very small percentage of the variation in student evaluations in the NSS is attributable either to universities or to subject areas within them. Most variation (about 90%) is attributable to the individual student. That is, the variation in the way different students evaluate the same course masks any average difference in the evaluation of different courses.

Students do not constitute a homogenous group, and some may value different aspects of their degree experience – be that face-to-face contact hours, strong employability focus, proximity to the best teachers, library facilities, etc. Such factors may also vary between disciplines. A degree of flexibility in any metrics used to assess teaching quality must be present, and considerable weight should be given to the institution's own written submission.

Students revise their opinions about the relevance and quality of different components of their degree after a year or two in employment, often coming to see courses in their degree, such as methods training, for which they may have shown little interest or enthusiasm at the time, as the most relevant for their careers.

The technical consultation argues that the new model for the subject-level TEF will be able to recognise and support diverse and innovative forms of excellence in teaching and provision. The metrics it proposes, however, do little to capture that diversity.

There is a danger that the use of student-satisfaction based metrics will discourage innovation in and provision of modules within degree programmes, for example on quantitative skills, if such modules, regardless of their pedagogical value, receive lower than average satisfaction ratings. Considering the evidence of a unique anxiety and aversion among many students in higher education when dealing with mathematics and statistics on their course, as identified in the Smith Review of post-16 mathematics, it is possible that essential moves to increase the level of quantitative skills learning across different subjects could lower student satisfaction ratings, despite such developments being crucial in improving the quantitative skills of graduates.

The ONS conducted a technical review of the NSS that reaches a similar conclusion to Marsh and Cheng. They show that almost all the 95% confidence intervals for student ratings of courses within each individual university contain the mean for all universities: that is, most university performances are statistically indistinguishable. This is true when averaging the

[1]Herbert W. Marsh and Jacqueline Cheng (2008) 'National Student Survey of Teaching in UK Universities: Dimensionality, Multilevel Structure, and Differentiation at the Level of University and Discipline: Preliminary Results', Dept. of Education, University of Oxford.

metric for all students at a provider, regardless of course. If more precise comparisons are made – e.g. by gender or ethnicity or social class of students, or by subject of study – confidence intervals become much wider and virtually no sensible distinctions can be made.

Among the core metrics that the consultation proposes to use, a measure of student outcomes – i.e. employment metrics – occupies a prominent place. It is not at all clear that the majority of the variance in employment outcomes is explained by tangible differences in the quality of different courses or the student experience that the provider offers. Rather, the prior attainment of students, their socio-economic background, gender and ethnicity, the institution's reputation and regional location, government macroeconomic policy, and the general performance of the economy, are likely to explain a much larger portion of the variance in graduate employment outcomes.

Data quality is also an issue. The Institute for Fiscal Studies research by Britton *et al.* (2016), working with a 10% sample of administrative data from HMRC found that data volume restricted what conclusions might be drawn. HMRC data has been used to create the LEO dataset and its limitations have been commented on by several sources. Useful and robust graduate employment outcome data would require timely and comprehensive data from the relevant government departments.[2]

The introduction of LEO data provided a useful additional insight into how graduate outcomes differ between subjects and institutions. However, graduate outcomes are not limited to employment and earnings. LEO data does not provide information on the types of role in which graduates are working, or how their higher education course might have equipped them to do it. In its recently published report, *The Right Skills: Celebrating skills in the arts, humanities and social sciences*[3], the British Academy articulated the skills that the 1.25m students who study arts, humanities and social sciences (AHSS) in higher education develop through their degrees. The report also investigates the contribution that AHSS graduates make to society and the economy. It showed that employment or earnings do not exhaust the contribution of graduates to society, which may come about through voluntary work or many other forms of public service.

Q8. Do you agree that grade inflation should only apply in the provider-level metrics?

The Academy is unconvinced that a reliable measure of grade inflation is achievable, either at provider level or at subject level. The difficulties lie in the ability to identify how much of between-year variance is the result of grade inflation or the result of other factors which affect grades, such as changes to assessments, fluctuations in the cohorts of students, or improvements in teaching.

A successful measure of grade inflation would have to show that the same cohort of students given the same assessment or series of assessments in year x as in year x-1 would receive a statistically different (higher) mark in year x compared with year x-1. It will be virtually impossible to accurately isolate this effect by controlling for all other relevant individual- and institution-level effects.

[2]Britton, J.; Dearden, L.; Shephard, N. & Vignoles, A. (2016) How English domiciled graduate earnings vary with gender, institution attended, subject and socioeconomic background, IFS Working Paper W16/06.
[3] *Ibid*.

Since one of the TEF's core aims is to improve teaching standards, and one consequence of improved teaching standards would be improved attainment of students, it is unclear how the Government intends to reliably differentiate between improved grades as a result of grade inflation or as a result of improvements in teaching quality. This renders the exercise problematic if not counter-productive.

The Academy urges the Government to consider carefully the reliability and efficacy of such a measure in light of these concerns.

Q10. To address the issue of non-reportable metrics:

a) do you agree with the proposed approach?

We agree with the Government's proposed approach to non-reportable metrics.

In view of the reliability issues with the core quantitative metrics, which we have covered elsewhere in this response, it is important that providers will be able to decide whether or not their metrics should be submitted. This flexibility is needed for the exercise to be valued and respected.

Q11. Do you agree that QAA Subject Benchmark Statements and PSRB (professional, statutory and regulatory body) accreditation or recognition should remain as a voluntary declaration?

Yes.

Learned Societies and Subject Association representing humanities and social sciences disciplines have shared their concerns regarding the reporting of PSRB accreditation to the TEF panels. Accreditation by a PSRB is far more common in STEM or Health than in the humanities and social sciences. Professional recognition is also more often likely to be found in more vocational courses than in degrees that facilitate the wider range of careers and employment outcomes, as is the case in the humanities and social sciences.

As accreditation is not relevant to all disciplines, the British Academy is concerned that institutions may concentrate on courses with professional recognition at the expense of others where such accreditations are less relevant if they believe that this will advantage them in the outcomes of the TEF.

Q12. Do you agree with our approach to capturing interdisciplinary provision (joint and multi-subject combined sources)?

The Academy believes that the proposed approach for joint programmes (i.e. counting each student pro rata against each subject in the subject-level metrics) is sensible.

However, the British Academy is concerned that the Government's understanding of, and approach to assessing, interdisciplinary study is problematic and could prove misleading to providers and to prospective students interested in interdisciplinary or multi-subject programmes.

Interdisciplinary programmes, in which academics draw from several fields and use methods across subject boundaries, go beyond standard disciplinary teaching expectations.

Interdisciplinary teaching does not simply replicate aspects of teaching and learning in multiple disciplines.

The approach suggested by the Government, which consists of using three general subjects for assessments and ratings, however broad they might be, will fail to capture the diversity and complex nature of the teaching provision in place in the vast majority of interdisciplinary programmes. This may have the unintended consequence of disincentivising providers from adopting interdisciplinary approaches in the future. This is particularly worrying if one considers how many of the key future global challenges will be interdisciplinary in nature and will require the combination of knowledge and skills across disciplines.

What the Government is proposing to do with that approach may well be relevant for multi-disciplinary programmes but meaningless for programmes that aim to be interdisciplinary. The Government must acknowledge the difference between joint- and multi-disciplinary study and interdisciplinary study and must provide a specific mechanism whereby the particular characteristics of interdisciplinary and teaching learning are examined and evaluated for quality and outcomes.[4]

Q13-15. 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?

The British Academy believes that the introduction of a measure of teaching intensity as set out in the consultation document will not contribute to the objectives of the subject-level TEF and might lead to unintended consequences.

All the options suggested in the consultation and technical document for the introduction of a teaching intensity measure present a large set of risks and disadvantages which, in our view, outweigh any benefits to measuring intensity.

The Gross Teaching Quotient (GTQ), which is the first option outlined by the Government, proposes to measure the contact time and weigh it by different measures including class size or the qualification and seniority of teachers.

While weighted measures of class size are more appropriate than non-weighted measures, linking teaching quality too closely to contact time will not be a reliable measure of quality in all disciplines. In his paper *Dimensions of Quality*,[5] Graham Gibbs demonstrated that the number of contact hours was unrelated to teaching excellence. He argued that what matters is what happens in those hours, not the quantity of them.

The British Academy believes that it would be better to focus instead on learning outcomes and experience. The new subject-level framework should compare how providers and departments use their resources to provide the best teaching and learning experience for their students.

There is a great diversity of teaching approaches in higher education, involving more and less contact time. Increased contact time may not be the route to independent thinking and

[4] British Academy (2016) *Crossing paths: interdisciplinarity, institutions, careers, education and applications*
[5] Gibbs, G. (2010), *Dimensions of Quality*, Higher Education Academy.

peer learning, which is a desirable outcome of any university teaching. As a measure, it might also fail to acknowledge alternative modes of study of learning based on reflexive and research-informed practice.

In addition, the contractual status, qualification or seniority level of staff involved in teaching is not a reliable indicator of the quality of their teaching. For example, research carried out for the Academy[6] demonstrated that postgraduate and postdoctoral teaching assistants for lab-work in quantitative skills proved to be of remarkable quality.

The second option outlined in the consultation document proposes to measure students' perception of the scheduled teaching received, as well as their own personal experience of the teaching. While the British Academy recognises the importance of students' views, it is very unlikely that surveys of students' perceptions will ever be able to account for and classify the diversity and pedagogy and teaching methods across providers and subjects. Moreover, there is a substantive risk that students' perceptions and personal experience of teaching received might not reflect the actual provision.

The next option listed by the Government requires collection and analysis of information submitted by providers on both expectations of number of taught hours and independent learning. Unfortunately, the document does not specify which types of modules and contact time the concept of 'taught hours' covers. There seems to be a case for the inclusion of varied tutorial modules, library inductions, and careers advice sessions, which, in turn, makes it difficult for such varied evidence to be aggregated and compared against at providers and subject level.

The complexities of data collection associated with the last two options outlined by the Government will constitute an additional and burdensome administrative process imposed on providers and departments without ensuring the robustness of these findings.

Higher education institutions are concerned with the advancement of knowledge and understanding, encouraging innovation and creativity in the way they teach their subjects. Ensuring that subject-level TEF is able to capture the diversity of teaching and learning experience and innovation at play in Universities is of utmost importance.

Q16. 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 Academy has not responded on questions relating to the two models for subject-level TEF because we believe that it is important for such a response to be based on evidence from the result of the subject-level pilots.

We are disappointed that the Government is consulting the sector about these models before it has been able to evaluate the results of the subject-level pilots. Although we appreciate the practical and time constraints, such a major new initiative should not be implemented without further consultation with the sector after the completion of the first year of the subject-level pilots.

[6] Measuring Up (2016). https://www.britac.ac.uk/publications/measuring.