



**Technical consultation on Teaching Excellence  
and Student Outcomes Framework at subject-level**

**Response from the Chartered Association of  
Business Schools**

11 May 2018

**The original consultation documents are available at:**

<https://consult.education.gov.uk/higher-education-reform/teaching-excellence-and-student-outcomes-framework/>

### ***About the Chartered Association of Business Schools***

The Chartered Association of Business Schools is the membership body of the UK's business schools and represents around 120 business schools and higher education providers, as well as affiliate stakeholders, corporate members and international partners. The UK's business and management education sector represents 1 in 7 university students and has a particularly high concentration of international students, accounting for 1 in 4 undergraduate international students and 1 in 3 postgraduate international students. It therefore has a strong interest in ensuring that subject-level TEF results in meaningful outcomes for both business schools and students, including international students.

#### ***Consultation questions***

***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?***

**Response:** Yes

**Feedback:** In principle we support the use of level 2 of the Common Aggregation Hierarchy to define 'subjects' in subject-level TEF as we believe that categorisations at this level will produce better quality information. In contrast, overly broad subject categories will obscure relative differences in performance between the constituent subjects and reduce the value of the resulting data for informing student degree choices.

***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?***

**Response:** Yes

**Feedback:** We do not agree with the use of the 7 subject groupings to conflate subjects within broad categories as this will not generate clear and coherent information for students - these are distinct disciplines in which students have very different educational experiences. The inclusion of Economics within the 'Social Sciences' subject group rather than 'Business and Law' is very problematic for some business schools, many of which have Economics tightly integrated into their overall course provision. The 7 groupings as currently articulated would require many business schools to make two submissions: one for Business and Law and another for Social Sciences. Feedback from the pilot exercise under Model B also suggests it is difficult to adjust figures (e.g. related to NSS) to remove data related to Economics courses.

It is also not appropriate for Law to be included alongside Business and Management. In many universities the subject of Law is taught in a completely different school, and feedback from the Model B pilot highlights the difficulties in working with a completely different group of colleagues in order to produce a joint submission. Furthermore, the way that a law school works may be very different from that of a business school.

The restriction of only 6 pages for subject submissions under Model B makes it very difficult to do justice to both fields, especially as the narrative is expected to show impact and quotes from students. Although the page limit is intended to save time, the experience under the Model B pilot shows the opposite, as a lot of effort is expended to stay within the word count, risking disagreements between the respective faculties as both want to use the very limited space to showcase their teaching performance.

In general the 7 subject groups do not align with typical institutional structures and whilst the proposal for Model B would enable institutions to move one subject in and out of each group this may not be flexible enough to enable all institutions to align their own structures with the groupings.

We believe that TEF at subject-level would generate more meaningful results for students if institutions were permitted to allow individual, per subject submissions for each of the 35 subjects based on the CAH2 definitions. We appreciate that this will have implications for the DfE in terms of resourcing, but we believe that to be useful to students subject-level TEF needs to be implemented at this level.

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

**Feedback:** Although a longer duration period would reduce the administrative burden of completing the TEF assessment, longer periods between assessments would result in outdated information. Older data will be of less value to students in addition to being potentially inaccurate given that even in a two year period there can be significant change to an institution's teaching methods and course provision. More recent information is obviously more useful for students.

In setting the reapplication period, consideration should be given to the importance of incentivising providers to make the changes in teaching provision required to improve their TEF ratings. A shorter re-application period would enable improvements resulting from such changes to be detected more quickly and enable providers to determine if they are on an upward trajectory. Providers with a provisional award lasting only one year should be permitted to re-apply the following year, but this would not be possible if the proposal for a two-year assessment cycle is chosen. Careful consideration is needed as to how subject-level TEF will accommodate providers with provisional awards.

A possible solution for accommodating improvements where a longer duration is chosen for subject-level TEF, would be for a provision for earlier reapplication where data from the metrics supports a different result. This would accommodate early recognition and better outcomes without being overly burdensome.

**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)?**

**Response:** Yes

**Feedback:** For the sake of consistency there should be alignment between the various TEF measures, so that the key elements at subject level are broadly the same as those at provider level, but some additional subject-specific information could be included e.g. professional accreditation.

**4) For the design of subject-level TEF, should the Government adopt:**

- a) a 'by exception' approach (i.e. a form of Model A)
- b) a 'bottom up' approach (i.e. a form of Model B)
- c) an alternative approach

**Response:** An alternative approach – we would support a revised Model B where assessment occurred at the level of the 35 subjects as opposed to being combined into the 7 subject groupings.

**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?**

**Response:** Yes

**b) allowing providers to select a small number of additional subjects?**

**Response:** Yes

**6) In Model A, should the subject ratings influence the provider rating?**

**Response:** No

**Feedback:** Allowing the subject ratings under model A to influence the provider rating would add further complexity to the assessment process and as 'exception' subjects are by definition exceptions to the provider-level rating it is unclear why it is proposed that they might influence the provider-level rating. We would caution against a version of Model A under which the subject ratings influenced the provider rating, with the exception of situations where a significant proportion of subject ratings diverged from the provider's overall rating. For such scenarios a specific threshold could be set which, if triggered, would prompt a re-consideration of the provider-level rating.

The technical document that accompanies the consultation discusses the merits of different approaches to identifying exception subjects. It notes that identifying exception subjects based on the criterion of subjects whose metrics differ from the provider rating by at least one flag would result in 54% of subjects across all providers being considered as 'exceptions'. Using a system based on subjects whose metrics differ from the provider-level metrics by at least 2 flags would produce a lower number of 'exceptions'. We would caution against a version of Model A that would make it overly difficult for a subject to be reviewed as this could result in students receiving a misleading impression as to the actual quality of a provider's teaching for that subject.

**7) In Model B, do you agree with the method for how the subject ratings inform the provider-level rating?**

**Response:** Unsure

**Feedback:** We question if the proposals for Model B could introduce a degree of subjectivity given that the consultation document talks about considering the different forms of evidence 'holistically' to determine the overall provider-level rating.

**8) Do you agree that grade inflation should only apply in the provider-level metrics?**

**Response:** Yes

**Feedback:** Measurements of grade inflation at subject-level may be greatly impacted by short-term volatility in subject-level grade movements which can occur for legitimate reasons rather than 'grade inflation'. By examining at institution-level a more indicative picture can be gained.

**9) What are your views on how we are approaching potential differences in the distribution of subject ratings?**

**Feedback:** The subject ratings should be allowed to vary naturally between subjects, rather than be forced into some kind of normal distribution which would be inappropriate as it could result in misleading comparisons regarding the 'value' of one subject relative to another. If TEF is to be useful

to students, it must give clear information on where the best teaching and student outcomes can be found for a particular subject. We concur with the technical document that forcing the ratings into a uniform distribution based on quotas would not represent an evidence-based assessment and could lead to a provider being awarded a rating that does not correspond to their actual teaching performance.

**10) To address the issue of non-reportable metrics:**

**a) do you agree with the proposed approach?**

**Response:** Unsure

**Feedback:** When attempting to deal with non-reportable metrics we would advise against using statistics based on small sample sizes which can be unreliable and misleading, especially if a TEF rating will have a duration of several years.

**b) when assessment occurs, do you prefer that assessors:**

**Response:** b) rely on provider metrics alongside any reportable subject-level metrics

**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?**

**Response:** Yes

**b) think that there are any subjects where mandatory declaration should apply?**

**Response:** Yes

**Feedback:** Mandatory declaration is only appropriate for programmes which offer professional accreditation or exemptions (e.g. nursing, accounting, law, actuarial science, medicine).

**12) Do you agree with our approach to capturing interdisciplinary provision (in particular, joint and multi-subject combined courses)?**

**Response:** Unsure

**Feedback:** The design of TEF should not discourage providers from adopting an interdisciplinary approach in their course provision. Generating metrics for joint and multi-subject courses needs careful thought because rating each joint degree individually will produce many non-reportable metrics and rating the constituent subjects separately will obscure the fact that they are quite different from the simple sum of the single honours degrees in those courses. Multi-subject programmes may be more beneficial than single subjects e.g. French may be less developed in terms of employment figures, but Business and French may increase employment opportunities. This value added may not be captured by using single subject ratings. Consideration should be given to awarding each joint and multi-subject course a specific rating if a credible methodology can be developed.

**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?**

**Response:** No

**Feedback:** The teaching intensity exercise included in the TEF pilot proved to be overly complex and time-consuming to complete. University systems are not geared towards reporting the data and pilot participants applied different understandings of the criteria in their returns. The information obtained from the teaching intensity measure is therefore of questionable value and arguably it measures the readily observable rather than the valuable.

While the fact that this data is not currently collected is not a reason in itself to exclude it from TEF, the lack of any common understanding of what constitutes 'teaching intensity', and its relation to teaching excellence and student outcomes, is even less understood, and for this reason should be omitted.

Neither contact hours nor class size are measures of teaching quality. It is not necessarily the case that more is better in relation to teaching intensity – the most important factor is how a class is taught. Teaching intensity metrics can be subject to temporary factors that might be beyond the control of the institution e.g. a period of high staff turnover. Independent learning is a key element of HE education and employers value evidence of graduates' ability to work independently, yet independent learning and blended learning are very difficult to capture through any teaching intensity measure.

Furthermore, any measure of teaching intensity would certainly need benchmarking at subject level as what is appropriate and feasible in one subject will not be in another. Further challenges to reporting teaching intensity are also posed by drastic variations in group sizes within units, events with no staff attached on the timetabling system, units with multiple staff attached (due to shared teaching responsibilities) and instances where a single unit is offered to students from different courses.

A focus on teaching intensity as a key measure will result in HEIs prioritising quantity of teaching hours rather than quality. If class sizes were a part of the measure of intensity and categories/thresholds of class size were used then it might lead to gaming the system by some institutions. There is also a risk that anything that isn't included in the teaching intensity metric will be cut from programmes in order to maximise performance under the metric.

***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?***

One of the many problems of developing a teaching intensity measure is the issue of what should and should not be included as there are so many measures that could be used across different subject areas, and there is clearly no 'one size fits all' approach. The following forms of contact and learning would lend themselves more easily to a measure of teaching intensity but would still prove problematic to standardise across so many different providers and subject areas: lectures, seminars, one-to-one meetings, online provision, field trips and blended learning. Information relating to the following would not be useful: student perceptions, independent study and perceived quality of teacher. A narrow focus on measures of teaching intensity has the potential to stifle pedagogic creativity and multi-modal delivery. This is incongruent with student feedback, employer work practices and contemporary societal trends.

***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.***

Seniority of teacher is a very inappropriate measure for when trying to determine teaching excellence. The pedagogic approach to teaching is more important than categorisations such as

lectures/seminars/work-based learning, and pedagogic approach is not easily categorised or measured. Genuine observations of teaching and student-teacher interactions are likely to be much more reflective of actual quality than simply quantifying methods of teaching delivery.

***16) Do you have any other comments on the design of subject-level TEF that are not captured in your responses to the preceding questions?***

On the plus side subject-level TEF forces institutions to reflect on granular data and performance of subject areas and to make strategic decisions based on analysis. It enables pockets of excellence to shine externally and thus provides greater transparency in the market to students but also exposes areas of strength and weakness internally. On the down side, it yet again will promote a certain amount of game playing and metric chasing and raises the question as to whether the overarching goal of employability - often measured by 'graduate outcomes' and salary - is the totality of a university education. By implication, does this really mean that someone working in a bank in London is considered to be successful compared to a social worker in Dudley?

The overall quality of the institutional setting is a strong factor in influencing the general quality of the teaching experience and this would be more effectively captured under the Model A approach. There are good reasons for retaining the approach of assessing HEIs holistically because student experience and graduate outcomes can be strongly impacted by factors at HEI rather than subject level. If Model B is chosen in its form as currently proposed, assessors will inevitably have to rely more heavily on the metrics rather than the subject group submission. It is not acceptable to compromise the quality of submissions simply in order to save resources.

In the case of the assessment of single subject providers under Model B, we note the intention that the provider-level assessors and the main panel will review the ratings fully, and that the agreed upon rating will serve as both the provider's subject-level rating and provider-level rating. We are concerned that this could mean that providers submitting only one subject will have this submission assessed more stringently than the subjects of larger institutions. This may put smaller universities at a disadvantage and we would like to see assurance that this would not be the case.

The proposed transfer of metrics from provider-level TEF to subject-level TEF is also an issue given the deficiencies in some of the metrics already used. The graduate salary metric is crude as it does not take into account median salary of the location of the university and uses a national figure. The range of graduate salaries in different parts of the country vary widely and so this measure disadvantages universities in relatively poor parts of the country, particularly important if the student population in the university is tied to a geographical area. The narrow focus on graduate salaries loses sight of the fact that roles in some industries pay significantly less than in other industries yet arguably generate a much greater social benefit e.g. social work versus finance.

NSS data are vulnerable to effects from matters beyond the control of subject providing schools and departments. The employment data are highly volatile, not only in ways that affect all subject providers equally (e.g. business cycles), but in terms of regional effects which are not reflected in the way that the core metrics are reported and used. Newly introduced programmes always have worse employment profiles than established ones: this may act as a disincentive to innovation in course provision. There should also be thought as to how the NSS can be improved, perhaps more in-depth opinions from students could be captured through focus groups.

We ask that more thought is given as to the value of TEF at subject-level from the perspective of the end-user – students – and how they are likely to perceive and act upon the different ratings of gold, silver and bronze. The consensus amongst our members is that most students are likely to draw the

simple conclusion that gold is better than silver and that bronze is an indicator of poor teaching quality. This could lead to distorted perceptions about quality of teaching which are not borne out by reality. In particular it has the potential to damage the international recruitment market for business schools, which accounted for one in four of all international students recruited to UK universities in 2016/17.

There is a risk that such misperceptions are further compounded by the proposals for Model B, under which it is theoretically possible for a provider to receive a subject-level based initial hypothesis of bronze even if 66% of its subjects are rated as gold. We would therefore suggest that students are closely involved in the development of TEF at subject-level to help ensure that the final system can be used to make meaningful and accurate decisions and does not lead to confusion and unfounded judgements.

For the full roll-out of TEF at subject-level we strongly urge that consideration is given to the most appropriate time to hold the assessments. The timing of the pilot exercise, which was held mid-way through the academic year, was not appropriate as it coincided with a period of heavy marking between semesters A and B, resulting in further administrative burden for those involved in the pilot.