

UNIVERSITIES UK RESPONSE TO SUBJECT-LEVEL TEF TECHNICAL CONSULTATION

This consultation along with the subject level pilots provides an opportunity to explore the issues and feasibility of a subject level TEF. In responding to this consultation, the sector is keen to help government and OfS explore these challenges and issues. This response does not, however, represent a sector endorsement of a subject level TEF at this stage. Significant questions over the implementation and added value of subject level TEF remain.

We would welcome further consideration about how the complex design challenges of subject level TEF could be resolved before taking a view on whether it can or should be implemented in full. Based on the proposals set out in the consultation and feedback from the pilots to date we are concerned that a viable subject level TEF would require substantial changes to either of the current options.

Given these ongoing concerns the future development of subject level TEF will need to consider the following questions:

- Whether it is viable for TEF to present robust and useful information to students using a benchmarked metrics led assessment model or whether a revised approach is necessary.
- Whether subject level TEF adds sufficient value to the information that is already provided to students about the structure and outcomes of courses and whether there may be better ways of supporting student decision-making
- Whether there are other ways that institutions can be supported to present robust comparable data on the impact of their teaching alongside a core set of metrics to inform good quality panel judgements.

We would welcome further discussion and engagement on these issues and challenges and it is essential that the independent review, a commitment set out by government to Parliament, is given a full opportunity to consider the future development of the TEF and the role, if any, that subject level assessment should play in this.

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?**

No - disagree

We remain concerned that this level of subject granularity undermines the coverage and quality of core institutional metrics. This will result in a significant reliance on panel judgements to different extents for different providers and courses. This raises concerns about the viability and consistency of using core metrics to form judgements about course quality when using this model.

Universities UK recognises the reason why level 2 of the Common Aggregation Hierarchy has been chosen as the preferred subject classification and that any approach will inevitably be a compromise based on the desire to provide detailed information to students, the burden of the exercise and availability of information to inform it.

The consultation document already considers two key criteria against which a possible subject classification method should be assessed: a manageable assessment process and an existing system. Whilst CAH2 meets the latter of these criteria and its ability to deliver the former is being tested by pilot activity, we have two further criteria against which we believe suitability of a subject classification system should be judged.

The first of these is its ability to provide robust data against which meaningful judgements can be made. We are concerned that the chosen CAH2 classification does not allow this. The number of subject instances with non-reportable metrics is large with 87% of providers having at least one subject with non-reportable metrics and 27 subject areas having at least 20% of providers with a non-reportable core metric(s). The consultation document highlights that 98% of students are covered within reportable metrics but absolute coverage is not the primary concern when the aim of the TEF is to inform student choice and to recognise and reward excellent teaching. To meet these aims all subject instances within institutions, regardless of current size, must be assessed on an equal footing. We do not believe that a combination of the chosen subject classification and a metrics-led assessment process allows for robust and comparable judgments across a sufficient number of subject instances.

The second criteria we believe to be important is whether the chosen subject classification adequately represents institutional structure or whether it is likely to limit institutional autonomy, leading to convergence and undermining diversity

within the sector. This would clearly not be in the best interests of students. UUK members who participated in the subject-level pilots were asked to provide feedback on their experience to inform this consultation response. Of those who responded (20 member institutions), none felt that the CAH2 classification represented their subject mix well, with 95% feeling it only somewhat reflected their subject profile.

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?

Yes - agree

When asked what they would change in order to improve subject-level TEF 39% of member institutions who have participated in the subject-level pilots felt a more granular subject classification was needed whilst none felt a higher level subject classification was required. We believe that feedback from the subject-level pilots and testing with students is the best way to assess what specific changes or tweaks may be required. Any tweaks considered should address the concerns set out above about decisions being based on robust and comparable evidence and the classification not impinging on institutional autonomy, leading to convergence and undermining diversity within the sector.

We are also concerned that the 7 subject groupings used in Model B, although designed to reduce burden, may have increased the complexity of compiling submissions. Feedback suggests that the groupings do not often align with institutional structures and therefore make it difficult to prepare coherent narratives and require the input of multiple stakeholders.

Member institutions who have participated in the pilots have also raised concerns about some courses being split across two or more subjects due to the way in which they have been assigned JACS/CAH2 codes even though they are not joint honours or combined courses. A major simplification would therefore be to only assign courses to their majority CAH2 code except in the case of a genuine joint honours programme.

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

The focus of this question is on whether we should extend the duration. However, please also provide as much detail as you can on your preferred length for the duration and/or re-application period.

Neither agree nor disagree

Any change to the duration of TEF award and re-application period must consider how best to balance the burden of the assessment process with the relevance and value of the judgments to students. Both of the options set out are likely to mean that students will be basing decisions on significantly outdated information that risks misleading students about what they can expect from the teaching and learning experience at the institution.

Consideration may need to be given to a hybrid methodology for re-application and award duration which takes account of any significant shift in an institution's performance against its benchmarked metrics over time. One possible manifestation of this approach could be a two-year assessment process plus an increased award duration but with mandatory re-application if provider-level metrics change sufficiently to lead to a different initial hypothesis.

Q3. 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)?

No – disagree

Universities UK recognises that subject-level information may be important for student decision-making but remains unconvinced that subject-level TEF is viable in its proposed form. The principle concern remains whether it will provide meaningful judgments that are based on robust information. Poor data coverage and the consequent implications for the consistency judgements risks misleading students.

In addition, there are also ongoing concerns about the likely impact on institutional operations, such as the homogenisation of subject structures, that do not improve the quality of teaching for students. Of our members who participated in the pilot activity only 25% of those who piloted either Model A, Model B or both felt that that model was suitable to implement sector-wide. Only 21% of member pilot institutions felt that model(s) of subject-level TEF they piloted would provide useful information to students or that it is based on accurate information about the teaching experience and outcomes at their institution.

Additional feedback from our survey of member pilot institutions, UUK's Review of Year 2 of the TEF and our ongoing work and member engagement on the TEF suggest that the existing framework has several design features which limit its ability to meet its current aims at provider-level. These issues are then compounded by extension of it to subject-level.

Concerns include:

- The use of a metrics-led approach that is based on poor or incomplete data
- Incentives that may undermine diversity and innovation in programme design
- Compounds issues with lack of alignment with the devolved administrations

With respect to whether subject-level TEF should retain the key aspects of the provider-level framework, we believe that the number of subject instances with non-reportable metrics necessitates a shift in assessment framework approach from one that is currently metrics-led to one that is metrics-informed. The ten TEF criteria seem broadly representative of teaching excellence and good student outcomes and could provide the basis for a reimagined framework. We also believe that benchmarking and independent panel assessment are key features of the current framework which should be retained.

Universities UK recommends that:

- the results of the subject level TEF pilot are considered fully as part of the independent review with an appropriate consideration of recommendations and the likely need for a fundamental revision if subject level TEF is considered. This should consider whether subject level TEF judgements provide useful information to students or whether students would value more structured information about the design of prospective courses.
- further consideration is given to the limitations of the trade-off between subject granularity, contextualisation and robust, high-coverage data. This should consider how subject level data may be incorporated in a metrics informed assessment process that enables comparable judgements of how an institution is delivering teaching excellence.
- any metrics used in the assessment framework are student-centred and better reflect the impact on students and their outcomes rather than the current focus on employment outcomes.
- the ten criteria, independent panel assessment and benchmarking/contextualisation of institutional environment are key features that, subject to the findings of the independent review, could be retained in any new framework.

Q4. 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)?**

When answering this question, please consider the underlying principles that define Model A (a ‘by exception’ approach) versus model B (a ‘bottom up’ approach), and which principle you think we should adopt for subject-level TEF. Whilst we are also interested in detailed comments on the specific design of each model, the final design will likely be a refined version of those presented in this document. This question is therefore seeking views about which underlying approach you prefer.

In your response, you may wish to consider the evaluation criteria set out in the specification for the first year of pilots. These are:

- Meaningfulness for students – the ability of the models to generate subject-level ratings that are meaningful for students and are more useful than the outputs of provider-level TEF.***
- Value for money – the proportionality of cost of participation for providers and cost of delivery for Government.***
- Robust processes and metrics – how well the models allow assessors/panels to make robust assessments, including how the metrics and submissions are used.***
- Supporting diversity of provision – the capability of the models to recognise diverse and innovative forms of excellence.***
- Effects on provider behaviour – how the models incentivise focus on and improvements to learning and teaching relative to provider-level TEF, and the extent to which the models avoid driving unintended consequences and minimise vulnerability to gaming.***
- Supporting widening participation and social mobility – how the models encourage providers to deliver positive outcomes for students from all backgrounds.***

As set out in question 3 we do not believe that either model represents a viable approach for subject-level TEF.

Feedback received from our members who have participated in the subject-level TEF pilots support this position. Across both Models only 25% of participants felt that the model(s) they piloted were suitable to be implemented sector-wide, with 33% believing the model to be unsuitable and 42% unsure for those piloting Model A and 50% thinking Model B unsuitable and with 25% unsure. When asked to expand on this, Model A participants felt that the method for generating exceptions was unclear and too complex, they also doubted how well a “by exception” model could inform student choice. Model B participants felt that the subject grouping process and the prevalence of non-reportable metrics were the biggest issue with the “bottom up”

approach as currently designed. Both of these issues led to participant concern about the extent to which subject-level TEF judgements resulting from this model would be based on accurate information about the teaching experience and outcomes at their institution.

Removal of the 7 subject groupings in Model B or at least increasing their flexibility would alleviate some provider concerns related to this model. We do have significant concerns, however, about the extent to which any metrics-led model will be able to deliver a robust assessment process or incentivise positive provider behaviour in either a “by exception” or “bottom-up” approach. This lends further support to our recommendation that the assessment framework move from a metrics-led to a metrics-informed approach.

Q5. 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?

Yes – agree

It is important that the pilot evaluation activity explores how the existence of clustered metrics and prevalence of high and low absolute values in certain subject areas affects the likelihood of certain subjects being raised as an exception.

b) allowing providers to select a small number of additional subjects? You may wish to comment on the options for identifying these. You may wish to comment on any variations or options we have not mentioned.

Yes - Agree.

Q6. In Model A, should the subject ratings influence the provider rating? Please provide as much detail as you can on why and how this relationship should be brought about.

No – strongly disagree

The “by exception” model is based on the premise that only those subjects differing from the provider-level assessment should be considered. If the pilots subsequently find that a large number of institutions have a significant number of subject awards which differ from their provider-level award this should be taken as an indication that the method of generating the initial selection of exception subjects is not fit for purpose.

There are also additional complications with allowing exception subject awards to influence the provider level award in Model A which make such a step undesirable. These include whether any weighting should be applied to the exception subjects (such as by headcount as is the case in Model B) and how the proportion of subjects generated as exceptions will be accounted for.

Q7. In Model B, do you agree with the method for how the subject ratings inform the provider level rating?

You may wish to comment on the method for calculating the subject-based initial hypothesis, as well as how it is used in the assessment process. We also welcome alternative approaches that do not use a subject-based initial hypothesis.

Neither agree nor disagree

Universities UK broadly agrees with the method for how the subject ratings inform the provider level rating in model B. We think it is right that subject-level assessment takes place first and separately and then is combined to come to a provider-level rating. We do, however, have concerns with certain aspects of this process, these are set out below:

- a) Some concern has been expressed about the transparency of provider-level judgments in the current assessment framework, this model will compound this issue by combining the judgments of up to 7 different subject panels with that of the provider panel. There will also be added subjectivity around how the subject-based IH is combined with the provider-level assessment to form a final judgment.
- b) Whilst weighting subject-level awards by headcount to form a subject-based initial hypothesis seems like the fairest way to account for the range and scale of provision offered by a provider we are concerned that it may result in unintended consequences. These may include provider's prioritising subjects with larger cohorts and providing an incentive to grow or close courses in certain subject areas based upon TEF outcome rather than student or employer demand. These consequences must be explored with providers participating in the pilots.
- c) If, as currently proposed, award distributions are allowed to vary across subject areas then we are concerned how this may affect provider level ratings and in turn provider behaviour who may then be incentivised or disincentivised from offering certain subjects,
- d) Referring to the combination of subject-level awards as a subject-based initial hypothesis is misleading. Full assessment has taken place and referring to them in this way equates them with provider-level initial hypotheses which are based on metrics alone.

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

If you are able, please provide information about how grade boundaries are set within institutions to inform whether our rationale applies consistently across the sector. Comments on the potential impacts of applying grade inflation only at provider-level are also welcome.

No –disagree

We have concerns about the current inclusion of a Grade Inflation metric in TEF assessments without further clarification about how this data is being interpreted by assessment panels and wider consideration of how this relates to conditions relevant registration conditions of the Office for Students, including B3 and B4. Until these two aspects have been clarified there is a risk that

- a) The assessment panel may misinterpret data or be inconsistent in their judgements about whether inflation or improvement is being observed in crude trend data that isn't disaggregated or contextualised
- b) There is a failure to join up approaches to grade inflation between the TEF, the OfS register and other factors such as incentives from league tables

UUK is committed to addressing the issue of grade inflation to ensure there is public confidence in the transparency and consistency of academic standards and to ensure that the degree classification system remains a useful tool for differentiating student attainment. The factors that affect the distribution of awards are multiple and include:

- a) Increased attainment at schools – so successful applicants have higher grades or more qualifications than they did previously
- b) Increased investment in student support, and corresponding increases in attainment for international students and for students with widening participation characteristics.
- c) Developments in teaching practices and investments in teaching
- d) Changes in student characteristics
- e) Subject/discipline
- f) Changing academic practices, including inflationary practice

UUK is committed to protecting the diversity and autonomy of the sector as a crucial feature of a successful higher education sector. UUK is also committed to criteria based assessment, based on learning outcomes rather than a norm referenced model that would demotivate most students.

UUK is currently working with the UK Standing Committee for Quality Assessment to review the issue of increasing proportion of 1sts and 2.1s. This work will help to improve the transparency and consistency of standards by:

- a) clarify sector reference points for the honours degree classification system
- b) develop analysis of the factors that are driving increases in proportion of good degrees
- c) make system wide recommendations to ensure the sustainability of the degree classification system, including its role in the TEF in the context of the wider regulatory framework

Q9. What are your views on how we are approaching potential differences in the distribution of subject ratings?

You may wish to comment on our approach to very high and low absolute values, clustered metrics and regulation by Professional, Statutory and Regulatory Bodies (PSRBs).

The issues highlighted here (particularly those of very high and very low absolute values and clustered metrics) add weight to Universities UK's belief that the current metrics-led assessment framework is not suitable for implementation at subject-level.

Neither option; either allowing award distributions to vary by subject or enforcing a certain distribution is across them, is desirable. The chosen option, of allowing the award distribution to vary across subjects, is undesirable because a provider's final award will then be impacted by its subject profile. A key feature of institutional-level TEF is the inclusion of subject profile as a benchmark criterion. This is crucial because it means that providers autonomy to decide which subjects they offer is not impinged and that they are not incentivised or de-incentivised to offer certain subjects.

The other option, standardising the distribution of awards across subject areas, is undesirable for the reasons set out in the consultation document, namely that it enforces artificial standardisation which will limit the accuracy with which awards can inform student choice.

We also do not believe that either of the options for dealing with very high or very low absolute values is satisfactory. Option 1, applying the same thresholds that are already defined for provider-level metrics will mean the number of institutions with very high or very low absolute flags will vary by subject area and for some combinations of subject areas and metrics, all institutions will receive a high or low absolute value flag. The reason this is unsatisfactory is because it affects subject areas in a non-uniform way which in turn may affect provider-level awards. In Model A it will also affect the likelihood of a subject being flagged as an exception. Option 2 is unsatisfactory because, as described in the consultation document, it will result in

some very high values in some subject areas being flagged as low absolute values. This is clearly undesirable as it could mislead students and would not recognise the high performance of some subject areas in delivering against the chosen metrics. Once again, this highlights that the current assessment framework, which relies on initial differentiation based on a set of metrics, is not appropriate for extension to subject-level.

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

a) do you agree with the proposed approach (see Figure 8)?

No – strongly disagree

Universities UK believes that the number of reported subject instances with non-reportable metrics and the proportion of them in certain subject areas is too high to simply amend the assessment framework in this way to deal with them. To do so will lead to a less robust assessment framework and concerns about comparability of judgments. As highlighted in our response to question one, whilst 98% of students may be captured in reportable metrics, the TEF's ability to inform student choice relies on the production of robust judgments across all subjects at all institutions that offer them. Therefore, the number of subject instances with non-reportable metrics is currently unacceptably high and must be reduced or the assessment framework changed to lessen the impact of this issue.

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)?**

Neither option is desirable and effort instead should be focussed on reducing the number of subject instances with non-reportable metrics or reshaping the assessment framework to decrease reliance on metrics.

Q11. Do you:

a) agree that QAA Subject Benchmark Statements and PSRB accreditation or recognition should remain as a voluntary declaration, if not, why?

Yes - agree

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

No, we can see no reason why inclusion should be mandatory in any subject. Institutions can still include this evidence in their submissions if they wish. Where employment outcomes vary in a subject on the basis of courses being PSRB accredited or not this will already be reflected in a provider's metrics. Where accreditation is important providers already highlight this information to prospective students to help inform their choice.

Q12. Do you agree with our approach to capturing interdisciplinary provision (in particular, joint and multi-subject combined courses)?
We want to ensure that providers are not discouraged from taking an interdisciplinary approach as an unintended consequence of subject-level TEF. We therefore welcome feedback on how the proposed approach will impact on providers and students.

Neither agree nor disagree

We believe this key area is best explored through the pilot activity but also that it will be crucial to investigate and understand the possible impact this treatment of these subjects will have on student choice. This is particularly important in the case of multi-subject programmes.

In addition to this we are concerned that these general subject areas have an extremely high proportion of providers with non-reportable metrics and that these subject instances are likely to be more prevalent in Scottish institutions. It will also be crucial therefore to understand the effect this treatment will have on institutional choice to participate across the devolved administrations. We believe that, in the interest of student choice, the TEF needs to encourage participation and ensure its fair across all countries in the UK.

Q13. 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?

No – strongly disagree

Universities UK is strongly opposed to the introduction of a teaching intensity measure in the TEF. Teaching intensity is not a measure of teaching excellence and we do not agree with the rationale set out for its inclusion in the consultation document. Nor do we believe there is any proportionate way in which teaching intensity, as described in the consultation document, could be measured which would not lead to burden on both students and providers allied to unintended consequences and incentivise perverse provider behaviour. These unintended consequences and

behaviours will be highly dependent on exact nature of the Teaching Intensity measure implemented but could range from homogenisation of teaching methods to maximising performance against the chosen measure by closure or withdrawal from subject areas with pedagogical methods not well-aligned with metric performance.

The rationale presented in the consultation document for the inclusion of a teaching measure in the TEF i.e. that it affects students perceptions of their studies, particularly their perception of value for money, that prospective students do not receive enough information on this matter and that contact hours and class size currently vary considerably across institutions and subject area, do not provide a persuasive argument for the inclusion of this measure in an assessment framework aimed at measuring teaching excellence and student outcomes.

Prospective students should have access to information about contact hours and class size when considering which institution to attend. However, that the TEF is not the correct way to provide this information. Contact hours should be presented as part of material information about programmes of study. Institutions should follow guidance on the presentation of this data to allow students to make a judgement on what is suitable for them.

We also do not believe there is any conclusive evidence that higher teaching intensity equates to better teaching or student outcomes and are concerned that its inclusion may have the opposite effect, stifling innovation, promoting a single pedagogical approach, worsening student outcomes. The consultation document points to findings detailed in Gibbs' dimension of quality which have shown class size to be negatively correlated with student achievement, engagement and depth of learning. However, on the subject of class size Gibbs' goes on to concludes that "in higher education what may matter most is not the size of the largest lecture that is attended on any particular course but the size of the smallest seminar group or problem class that they attend within the same course" highlighting that, because of the range of pedagogical approaches taken in higher education compared to other educational settings, smaller does not always equal better.

Gibbs' dimension of quality also considers contact hours as a measure of quality and states "the number of class contact hours has very little to do with educational quality, independently of what happens in those hours, what the pedagogical model is, and what the consequences are for the quantity and quality of independent study hours." It is clear therefore that teaching intensity is far too blunt a measurement to be able to capture teaching excellence with the "right amount" being highly context dependent and any relationship between it and teaching excellence being non-linear. In short, teaching intensity is not a measure of teaching excellence, teaching excellence is a consequence of good course design which will be subject and context dependent.

Feedback received from member institutions who have participated in the pilot activity and therefore also trialled various teaching intensity measures supports this position. No institution believed that it was appropriate to include a teaching intensity metric in the TEF, with 90% stating it was inappropriate and 10% unsure.

The introduction of any teaching intensity measure risks undermining institutional and academic autonomy to choose the correct pedagogical approach for their unique mix of subject and student cohort. It would lead to homogenisation across the sector, stifle innovation and damage student success.

Q14. 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?

We do not believe a contact hours metric is appropriate. Member institutions involved in pilots had a large number of widely held concerns about the teaching intensity collection method they piloted. 95% were concerned about student response rate, 90% about the quality of data collected, 85% about the comparability of data across institutions and 80% about the cost of collection.

Q15. 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.

If you have an alternative suggestion, you may wish to consider the following factors:

- *Meaningful for students – the ability of the measure/method to provide meaningful information for students.*
- *Value for money – proportionality of the cost of a measure.*
- *Generalisability across the sector – how a measure can be applied and work across the sector.*
- *Accuracy/validity of measures – how accurately data can be collected and verified.*
- *Supporting diversity of provision – the capability of the models to recognise diverse and innovative forms of excellence.*

We would strongly oppose the inclusion of any of the suggested teaching intensity measures in the TEF for the reasons set out in response to question 13. The comment section of Table 5 also adequately captures the individual deficiencies of each suggested method should it still be considered desirable to include a measure in the framework.

The TEF should focus on capturing student satisfaction with their educational experience rather than seeking to intervene in how institutions design their courses. Student reporting on the quantity of contact hours as suggested in the current student survey on contact hours cannot assist judgements of teaching excellence.

Q16. Do you have any comments on the design of subject-level TEF that are not captured in your response to the preceding questions in this consultation?

No.