Response to subject-level TEF technical consultation

1. Summary

1.1 We share the Government's ambition to provide meaningful information to prospective applicants and to promote the enhancement of teaching and learning. We are therefore concerned that the proposed design of subject-level TEF risks amplifying flaws in the methodology at provider-level and that, in turn, this has real potential to mislead prospective applicants and fail to provide a useful tool for enhancement.

1.2 The pilot exercise for subject-level TEF has been very useful in helping to identify significant problems with both of the models under consideration. Given the importance of providing useful information for prospective students, and the complexity of delivering subject-level TEF as currently constructed, we recommend that the pilot period should be extended further and that alternative models should also be considered for development.

1.3 We would like to work constructively with the Government to explore different approaches that could be taken in these pilots. As a starter, our experience suggests that any on-going development of subject-level TEF should ensure that:

   (a) results are methodologically and statistically sound, with high absolute performance recognised and rewarded automatically. Flawed teaching intensity and grade inflation metrics should also be removed; peer-reviewers should have sufficient subject-specific expertise to assess quality of teaching for each subject; and further guidance is needed to ensure decisions are consistent

   (b) information is sufficiently granular and comprehensive to be of use to prospective applicants. Subjects should not be aggregated together for the purposes of peer review (as in Model B) nor should they be exempted from assessment (as in Model A); a hybrid model where all subjects are assessed without aggregating them together for the purposes of peer review could provide prospective applicants with more helpful information. It would, however, add to the substantial burden of administering the exercise and, along with the other concerns set out in this paper, may therefore necessitate a considerable redesign of the proposed approach to delivering subject-level TEF. In addition, subject grouping "labels" need to be easily identifiable to students – a more granular grouping system than CAH2 may therefore be required

   (c) providers are treated equally, but with sufficient flexibility to avoid distorting the results. Some of the proposals advantage certain types of provider over others: for example, the calculation of submission length favours providers with a narrow range of subjects and should be amended to better reflect variations in size; and the treatment of providers with non-reportable metrics may lead to a "default" initial hypothesis of silver without supporting evidence to justify this. A minimum student cohort size per subject should therefore be introduced to be eligible for a full assessment and award

   (d) diversity of provision and innovation is incentivised, not penalised. For example, the proposed approach for inter-disciplinary and multi-programme courses would not provide students with useful information and may discourage institutions from offering such programmes, limiting student choice and stymying innovation

   (e) costs are minimised where possible and are proportionate to benefit. It is difficult to see how costs can be reduced without further sacrificing the validity of the results and the usefulness of the exercise for applicants. However, it would be helpful to explore ways of removing duplication with existing processes – including whether PSRB accreditation for certain subjects could remove the need for a separate assessment, and to what extent TEF submissions could be aligned with internal quality assurance procedures.

2. Overarching priorities

2.1 We welcome the opportunity to respond to the technical consultation on subject-level TEF and also look forward to engaging with the forthcoming independent review. We share the Government's desire to provide meaningful information at a more granular level to assist prospective applicants in their decision-making as well as to promote enhancement of teaching and learning and so improve teaching quality across UK HE. In developing TEF any further it will be important to consider whether and how these core objectives can be met, whilst avoiding an unduly burdensome exercise for Government, providers and students (not least because the costs of delivering TEF will at least partially be covered by their fees).

2.2 Overall, we are concerned that neither of the proposed approaches to implementing subject-level TEF would adequately address the core objectives of the exercise. We have a number of specific concerns regarding the rigour of the methodology which are likely to be exacerbated as it applies at subject-level:

- **Metrics**: the current metrics do not directly measure teaching quality, although this is one of the most important factors for students.[1] The inclusion of teaching intensity and grade inflation metrics, in particular, is a concern and risks undermining institutional diversity by encouraging homogeneity of provision. Student disengagement with the NSS appears to be a long-term issue and risks undermining the validity of the results.

- **Benchmarking**: a number of key flaws in the provider-level methodology are likely to be amplified at subject-level: small student numbers (particularly for splits) risk skewing outcomes; universities contribute to their own benchmarks making them harder to beat (especially for larger providers); and very high absolute performance may still go unrecognised (for example, where a provider does not score in the top 10% for a particular metric even though absolute performance is excellent, or where they have a negative flag). There also appear to be inconsistencies in the application of benchmarking by different metrics at subject-level.[2]

- **Generation of significant flags**: when forming the initial hypothesis all metrics are treated equally by simply counting flags. This takes no account of the degree to which the flags have either been met or not been met, so patchy extreme performance could be rewarded over consistent strength. The generation of flags as a binary process (you either have a flag or you do not) also creates an artificial and unjustified cliff edge.

- **Descriptors**: the gold/silver/bronze ratings are blunt instruments with harsh cliff edges between bandings and do not recognise the rigorous quality assurance system which institutions must satisfy. As a result, bronze and silver are likely to be interpreted as "not excellent" or worse, with potentially negative impacts for domestic and international recruitment. Using simplistic descriptors also fails to recognise the diversity of provision.

2.3 **Given the flaws in the methodology for creating the initial hypothesis, the peer review element of TEF is critically important in enabling institutions to provide a wider range of qualitative and quantitative evidence regarding their approach to teaching in each subject under assessment.** To ensure the quality of any peer review element, panel members should have an appropriate level of expertise with regard to the subject in question

---

[1] Recent research on value-for-money by the Office for Students found students believe quality of teaching to be the single most important factor demonstrating that an institution offers good value for money. https://www.officeforstudents.org.uk/new-research-shines-spotlight-on-student-perceptions-of-value-for-money/

[2] For example, the use of nine broad subjects for the benchmarking of DLHE data creates misleading outcomes. Architecture is benchmarked against social science subjects such as business and law, for instance, when the employment environment is very different for such disparate programmes.

(e.g. subject specific knowledge and understanding of the ways in which teaching is delivered and how students learn in their given field). Further guidance is needed to ensure panellists are supported to make consistent decisions at subject-level and it would be helpful for the pilots to explore whether a more structured common template for written submissions would encourage consistency of assessment.

2.4 **The current complex methodology is very difficult to understand and explain to students or other stakeholders, and risks distorting public interpretation of the results and undermining the usefulness of the exercise**. Research undertaken by a consortium of students' unions[3] found no evidence that students understand TEF ratings are based on benchmarking, and not absolute performance, rather they assumed it was valid to compare one gold institution with another. The research also found some evidence that students may not interpret TEF ratings as intended, with negative consequences for social mobility: a minority (6%) of students would reconsider applying to, or not have applied to, their current institution if it had been rated gold, and this number is higher for BME students (10%).

2.5 Because TEF results are difficult to interpret, there is a significant risk subject-level TEF could damage the reputation of UK HEIs overseas and so affect international student recruitment. Whilst it is still too early to know how TEF may impact on international applications, surveys show international students do not understand TEF results: a quarter think bronze indicates teaching quality is unsatisfactory.[4] As part of the research being undertaken with students alongside the consultation, the Government should seek to assess the extent to which applicants understand what TEF measures and test whether the information provided is meeting their needs.

2.6 **It is too early to tell whether TEF is having a positive effect on enhancement of teaching quality within institutions**. A survey conducted by UUK found that the TEF has only influenced around 20% of responding institutions to review and revise existing learning and teaching enhancement activities, whilst much of the additional investment in this area had been planned before the TEF was introduced.[5] The focus on competition over collaboration may limit the extent to which TEF can drive enhancement and some elements of the methodology do not encourage institutions to focus on the areas of performance which require the most improvement. For example, the use of flags as binary measures creates perverse incentives to focus on areas where institutions are almost meeting or are slightly above their benchmark rather than focusing on their weakest areas where improving on the benchmark would be more difficult.

2.7 **Any meaningful and robust subject-level exercise is likely to require a significant bureaucratic structure to support it**. UUK has estimated that HEIs spent £4.1 million on staff costs to participate in TEF2 and we can expect this to be multiplied substantially at subject-level, whilst the estimated cost to the Government of running provider-level TEF is £2.1m rising to £19m for the first year of subject-level TEF (although the real cost at subject-level is likely to be much higher as the exercise as currently constituted is more complex than that which informed previous estimates).[6] The total cost of delivering subject-level TEF is therefore likely to be substantial, and it is imperative that the value the exercise delivers should be proportionate. **Indeed, it is questionable whether a subject-level TEF would pass any normal value for money test based on experience from the pilots. There are**

---

[3] https://studentsunionresearch.files.wordpress.com/2017/11/tef-pr-research-report.pdf
[4] https://thepienews.com/news/hobsons-survey-reveals-most-international-students-confused-by-uk-tef/
[5] http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/review-of-the-teaching-excellence-framework-year-2.pdf
[6] *Detailed Impact Assessments: Higher Education and Research Bill*, June 2016

already several subject rankings available so subject-level TEF must add considerable value for prospective applicants to avoid needlessly duplicating existing information.

2.8 In addition, we can anticipate substantial opportunity costs to universities and the wider academic community of participation in subject-level TEF. Depending on the timing and the extent to which a rolling assessment is operated, there may also be competition between the REF and TEF for the best academics to participate in the peer review process, which could affect the ability of each exercise to attract the very best experts to undertake peer review functions.

2.9 **It will be difficult to mitigate all these risks.** We therefore welcome the intention to pilot subject-level TEF for a second year. Given the complexity in delivering subject-level TEF, **we recommend that the pilot period should be extended further** to allow time for any preferred model, or alternative options, to be sufficiently tested. Even then, it is difficult to see how subject-level TEF as currently envisaged could provide genuinely useful information for prospective students given the flaws with the underlying methodology. We therefore look forward to engaging with the independent review of TEF to consider how an exercise may be developed which adds significant value for prospective applicants, Government and providers.

2.10 For example, research has suggested that rather than focusing on the provision of additional information, for prospective students more attention should be paid to supporting the use of existing information more effectively, with better integration and clearer links between the various sources already available.[7] A better way of addressing the need for accessible and relevant information for applicants may therefore be to bring together existing data and information sources in a more sophisticated, real-time, and user-friendly online interface. Developing a new student information tool could also provide an opportunity for more personalised information recognising that students are by no means a homogenous group, and so help them to make sense of the information they're presented with and raise their aspirations.

2.11 We have set out our initial priorities for any further development of subject-level TEF in the following sections of this response and we intend to develop these in our evidence to the independent review. The pilot exercise has been very helpful in identifying opportunities and shortcomings in the different approaches being trialled and we look forward to working with Government to explore how TEF might develop in a way that is more useful to students and providers while reducing cost and burden.

3. Subject classification system

3.1 The use of the Common Aggregation Hierarchy Level 2 as the classification system for subject-level TEF means that distinct courses will be aggregated together which are different in their structure, design and approach to teaching, learning and research, but treated as though they are the same. Crucially, there does not appear to be any evidence that teaching quality across different courses in the same group will be "reasonably similar" as suggested in the consultation document.

3.2 Providers may therefore find it difficult to convey the individual performance and specific strengths of very different courses which must be presented collectively, and **the aggregation of lower-performing courses with highly-performing ones will lead to misleading results**. For prospective students, this could mean that the result for the most

---

[7] https://assets.publishing.service.gov.uk/media/53355970ed915d630e000017/OFT1529s.pdf

relevant subject grouping could be unrepresentative of the specific course they wish to undertake.

3.3 **It is also likely that some students will be unable to use results of the assessment as they may not even be able to identify the subject grouping which is relevant** to the course they are interested in studying. For example, will a student be clear about the courses and content within the "subjects allied to medicine" grouping compared with "health and social care", or in "engineering" compared with "technology"? It is crucial students are able to identify which subject grouping is relevant to them if subject-level TEF is to provide useful information[8]. It may be that a more granular version of the CAH2 system is needed to achieve this.

3.4 Furthermore, the intention to group the 35 subjects into seven groups for Model B would exacerbate the issues outlined above. These broad groupings are unlikely to align with institutions' own structures and provision, making it difficult to produce a coherent narrative about the performance of dozens of disparate courses. The proposed solution, to enable providers to move one subject in and out of each of the seven groups, is unlikely to be sufficient to enable institutions to present subjects according to the structures and teaching approaches within their university. Whilst greater flexibility in moving subjects in and out of the broader groupings could help, this would limit the extent to which comparable results could be generated which are helpful for students.

4. Models A and B

4.1 **As currently proposed, neither of the proposed models for delivering subject-level TEF are capable of generating helpful information for prospective applicants or offering sufficient opportunities for institutions to demonstrate excellent teaching.**

4.2 In summary:

- Model A is likely to mislead students about the way in which subjects have been assessed (as some will be fully assessed and others will not) and the process will be unpredictable for institutions and the Government to manage. It would also fail to provide sufficient opportunities to demonstrate excellence, for example only assessing subjects with silver and gold initial hypotheses where the provider rating is bronze would not provide the opportunity for those subjects with the most to gain (i.e. bronze subjects) to improve on their initial hypothesis.

- Whilst Model B would assess all subjects and so avoid the issues of inconsistency outlined above, it would be likely to place a very considerable burden on participating providers. As above, aggregating subjects together into seven broad groups for the assessment of submissions is unhelpful as the aggregation of lower-performing courses with highly-performing ones will lead to misleading results of limited use to prospective applicants. The assumption this approach to peer review would reduce some of the burden is ill-founded as it may in fact involve additional work for institutions where the starting point of crafting a submission is to consider performance at subject level.

4.3 The table below sets out a more detailed assessment of our concerns regarding each model:

---

[8] This is particularly the case where ostensibly similar subjects may be grouped differently (and therefore be rated against different benchmarks and standards) at different institutions.

Table 1: analysis of models A and B

For both models	Flaws with the benchmarking methodology at subject-level and issues with reportability of data are likely to undermine the rigour of the results Some students may also struggle to identify which subject-grouping is relevant to them (see the section on the Common Aggregation Hierarchy, above)	
	Model A ("by exception" approach)	**Model B ("bottom up" approach)**
Will the information be useful to prospective students?	• The assessment process is not consistent for every subject, but this is not reflected in the ratings and so could be misleading	• While this Model may appear to provide prospective students with more granular information, as with Model A the subject groupings may not be recognisable to students. • The complex methodology with metrics assessed at one level and submissions pitched at another level is also likely to confuse students.
Are results likely to be robust?	• It is unclear how the initial hypothesis (IH) will be used to identify "exception" subjects and how this might work for institutions where the provider-level rating is better than the IH based on the provider-level metrics. • The five-page submission per exception subject is unlikely to provide enough space to demonstrate excellence. • The way in which the exception subjects are chosen may also limit opportunities to demonstrate excellence. For example, if a provider receives an overall IH and rating of silver, then subjects with gold and bronze initial hypotheses will be assessed, but it would be unlikely for a gold subject to be downrated based on the submission; a more fruitful approach could be to consider whether any of the subjects with a silver IH merit uprating.	• We expect the flaws in the TEF methodology to be amplified at subject-level assessment, so the way in which this feeds into the overall provider-level rating under Model B is a concern. • Grouping subjects together for the assessment of submissions assumes that subjects often found in the same faculty or department are broadly comparable – which is not necessarily the case. We might expect more silver ratings as bronze and gold subjects balance each other out. • The length of submission for each broad group is unlikely to be sufficient to enable providers to demonstrate excellence across a range of subjects, and the approach to varying the length of submissions by number of subjects an institution provides favours those with a narrower range of subjects.
What is the expected administrative burden?	• Difficult to anticipate the burden given the number of submissions could vary widely (depending on the number of subjects to be assessed). • Whilst the intention is for subjects to be assessed "by exception", subjects may behave differently from the average because of the benchmarking methodology (e.g. benchmarks will be impossible to beat for courses such as medicine and dentistry). • Unpredictability in the number of subjects to be assessed will be challenging for providers, assessors and panels to plan for.	• Likely to place a very considerable burden on participating providers: they will need to produce up to 73 pages of submissions at provider- and subject-level. • This is in addition to the reporting burden imposed by the teaching intensity and grade inflation metrics, which is also considerable for both models.

	• Whilst this model is presented as the less burdensome option, the length of submission will still be very considerable: 15 pages for provider-level and 5 pages per subject.	

4.4 In addition to the issues outlined above, both Models A and B risk confusing prospective students by allocating ratings at subject-level which may conflict with the provider-level award. This will mean that students may be studying in a gold-rated provider but in a bronze-rated subject and vice versa, and it is unclear how they might be expected to use this information to inform their decision-making.

4.5 The relationship between the institution-level and subject-level outcomes for Model A is even more complex. Whilst the initial hypothesis at provider-level will be used to identify the exception subjects, this ignores the possibility that the final provider-level outcome may not match the initial hypothesis. If a decision is taken to continue to pilot Model A, the assessment of the provider-level and subject-level submissions could be staggered so that the provider-level submission be submitted and assessed, and the rating confirmed, before the exception subjects are identified.

4.6 Assessing subjects separately from performance at institution-level creates an artificial distinction. For example, for Model B, all ten TEF criteria are assessed at subject-level whilst only three (valuing teaching, resources and "positive outcomes for all") are assessed at institution-level. An alternative approach would be to utilise information on the learning environment at institution-level in much the same way as the "environment template" is used for REF submissions. This could involve applying some standard institutional text to all subject submissions followed by more detail on how this is translated at subject-level.

4.7 **Given the extensive flaws with both models A and B, we recommend that the Government should also pilot a hybrid model where all subjects are assessed (rather than exempting some as in Model A) and without aggregating them together for the purposes of peer review (as in Model B)**. Such a model could be more likely to provide prospective applicants with helpful information about subjects they are interested in studying and may offer greater opportunity for institutions to demonstrate excellence at subject-level. It would however add to the considerable burden of administering the exercise for the Government and providers, and would not address all of our concerns with Models A and B as outlined above, especially the way in which flaws in the TEF methodology are likely to be amplified at subject-level and would feed directly into the overall provider-level rating.

4.8 Indeed, a move to a hybrid model may necessitate a considerable redesign of the proposed approach to delivering subject-level TEF in order to control the financial and opportunity costs placed on institutions, and so avoid this impacting on the ability of institutions to deliver across their core missions.

5. Metrics

Teaching intensity

5.1 **There are substantial challenges associated with any measure of teaching intensity**. Crucially, it would not provide any insight into the quality of the contact hours students receive or the type of interaction the student has with the academic – and so would not meet the needs of prospective applicants whose top priority is to receive information about the

quality of teaching.[9] It is likely to be particularly difficult to capture independent learning reliably – a key feature of studying in a research-intensive learning environment and essential for students in developing and consolidating knowledge as well as facilitating wider development.[10]

5.2 Our key concerns with each of the models of teaching intensity outlined in the consultation document are as follows:

Table 2: key concerns regarding teaching intensity options

Model	Key concerns/issues
GTQ	• Takes no account of independent learning or quality of provision delivered • Potential to drive institutional behaviour towards greater homogeneity, for example, by placing greater value on small group teaching, even if this is not appropriate • Likely to impose a significant burden for institutions in reporting on the data, particularly as internal systems may not be set up to deliver this data
Student survey	• Any student survey as part of TEF may be vulnerable to disruptive action (as has been the case for NSS) – even if this does not happen, response rates are likely to be low as a result of "survey fatigue" compromising the efficacy of the results • Students' perceptions of their own engagement may not match reality and may differ depending on their own preferences and priorities
GTQ weighted by qualification/ seniority	• In addition to the issues outlined above, the key challenge of this approach is the lack of any recognised proxy for what a "good teacher" should look like • There is no evidence that the qualification and/or seniority of the teacher would be an adequate proxy for quality of the teaching • Individual institutions will each have their own conception of what qualifications are required to be a "good teacher" including their own bespoke professional development schemes. Applying a blanket definition could damage an institution's ability to innovate in this space and risk undermining institutional autonomy
Quantitative & qualitative information about how students are expected to spend time	• Whilst this approach avoids being prescriptive about what a good model of teaching and learning should look like, it is very likely to be open to gaming as providers could exaggerate declared levels of self-study • It would also introduce significant additional burden on institutions
A measure of engagement with teaching resources	• There are a wide range of practical issues associated with a measure of student engagement with teaching resources: many institutions will not measure student attendance, VLE activity, library visits etc, so new processes would need to be put in place; even where engagement is measured datasets may not be comparable • Crucially, would not provide insight into the quality of engagement or how students are supported by academic staff • Unclear how values would be assigned to different types of engagement and how this could be compared usefully across institutions; would require a complex methodology • It may also be problematic for providers to both store and use information on student engagement with campus facilities and other resources – consent may be required

---

[9] The recent OfS survey highlighted that students are as likely to associate poor value for money with poor quality contact time as they are with low levels of contact time. See footnote 1.

[10] Thomas, L., Jones, R. & Ottaway, J. (2015) *Effective practice in the design of directed independent learning opportunities.* York: HEA & QAA

Measure of staff contracted teaching hours	• Particularly problematic for institutions with a research-intensive learning environment given that staff are often contracted to engage in both teaching and research • There are also staff members who teach across different departments and courses and establishing how these hours should be classified would be challenging • Would not account for hours spent preparing for teaching or provide insight into teaching quality • More generally, this measure would represent an even blunter tool than the GTQ approach to considering teaching intensity and would likely be open to gaming

5.3   **Given the significant problems with any measure of teaching intensity, we recommend that it be removed from the TEF**. This would avoid the risk of skewing the pattern of teaching provision across providers in order to perform well against such a metric, which, in turn, we can foresee would lead to homogenisation and ultimately impact on the quality of education and student choice.

5.4   Alternatively, providing additional information about the teaching and learning environment within institutions may better meet the needs of prospective applicants. Recent research shows students want TEF to include data on facilities such as IT (86%), online resources, the library and learning spaces (93%) as well as subject specific resources (94%).[11]

Grade inflation

5.5   **Whilst the proposal not to use a "grade inflation" metric at subject-level is welcome, we remain concerned about its use within the provider-level exercise**. Such an approach fails to recognise that there are a number of factors which can affect trends in degree attainment including prior attainment, subject mix, student characteristics, improvements in teaching practice and student engagement, and so on. Indeed, if TEF is effective in improving teaching quality, we would expect attainment to rise over time. Trends in degree attainment represent a complex picture and make it difficult to isolate cause and effect. It is therefore unreasonable that the burden of proof to demonstrate that "grade inflation" has not taken place will reside with the provider, with assessors being advised to accept arguments only where there is clear and robust evidence to support them.

National Student Survey (NSS)

5.6   The continued boycott of the NSS at some institutions as well as concerns that results could be heavily skewed by external events (such as the recent strike action) suggest continuing to use NSS scores in a robust way to inform TEF outcomes may not be possible. There are pressing concerns about the validity of the NSS results for some institutions and how this will be reflected in future TEF exercises as well as how the results will be used more broadly to inform student decision-making. There will also be significant knock-on consequences for universities who interrogate the results to inform improvements in curriculum, teaching and learning quality, learning resources and academic support. **As a matter of urgency, DfE should seek to address student disengagement with the NSS and consider how the survey results will be treated to ensure institutions are not unfairly penalised.**

Non-reportable metrics

5.7   As currently constituted, the proposed methodology at subject-level is likely to lead to a significantly higher number of non-reportable metrics, especially at small providers. The

---

[11] https://studentsunionresearch.files.wordpress.com/2017/11/tef-pr-research-report.pdf

proposed solution (to lower the suitable metrics threshold and substitute non-reportable metrics for group- or provider-level metrics) is unlikely to provide useful information to prospective students as the initial hypothesis will not be based on the CAH2 aggregated subject grouping but on a far less granular grouping or even just on provider-level performance which may not reflect performance on a specific course. This is because the subject rating would likely be heavily skewed by performance in other subjects and the final subject rating could therefore be misleading.

5.8   Such an approach would mean providers with non-reportable metrics would be treated differently from those with the full suite of metrics, but both would still receive the same award. Enabling providers with very few metrics to be assessed based on aggregated metrics may simply lead to more initial hypotheses of "silver by default", as negative or positive flags become more unlikely. This would risk misleading prospective applicants about the performance of such subjects and could advantage smaller providers with fewer reportable metrics over larger providers where metrics are more likely to be reportable.

5.9   **A different approach may therefore be needed for very small providers to recognise that the current methodology cannot apply fairly where metrics become unreportable**. This could involve setting a minimum student cohort threshold to be eligible for a full assessment and award at subject-level.

5.10   **In addition, we remain concerned about how the low threshold for non-reportable metrics (at just ten students) is likely to impact further at subject-level on the rigour of the results**. Whilst raising the threshold would result in more non-reportable metrics, it would avoid the experience of very small groups of students skewing results and so undermining their usefulness to prospective applicants.

6.   Recognition of high and low absolute values

6.1   The consultation asks about how very high and low absolute values could be captured at subject-level. However, the way in which very high and low values are identified at provider-level (through stars and exclamation marks rather than flags) is complex and confusing for prospective applicants and this needs to be addressed at provider- and subject-level. The Government has already recognised that universities that have consistently maintained very high performance may not be able to deliver continuous and significant positive differences from benchmark values, and as performance on the metrics improves, the 2% substantive difference threshold test (materiality test) becomes harder, and sometimes practically impossible, to achieve.

6.2   **To recognise and incentivise high performance more effectively, a positive significance flag should be given automatically where a provider scores in the top 10%.** Building recognition of very high performance directly into the calculation of the initial hypothesis would send a clearer message to assessors and better recognise that the benchmarking methodology is flawed when applied to very high-scoring institutions. Such an approach would also improve the quality of the information generated by the exercise for prospective applicants recognising that students do not understand TEF ratings are based on benchmarking (see section 2).

6.3   Whilst it is essential that high performance is adequately recognised at subject-level, there are drawbacks to either option outlined in the consultation as to how thresholds should be set for each subject (i.e. applying the same thresholds to identify very high and low absolute values that are already defined for provider-level metrics versus applying different thresholds for each subject). As stated, the first option would avoid a situation where subjects such as medicine and dentistry would be penalised as a result of the very high benchmarks applied

to them, but there would then be less opportunity to recognise high absolute scores in other subject areas as a result. Further testing through the pilots is required to address this issue, in particular to ensure the flag/no flag cut point is not made on tiny differences in performance on metrics where providers are clustered.

7. Distribution of subject ratings

7.1 **We would not support a forced distribution of ratings across subjects as such an approach would limit the opportunity for improvements to be recognised year on year.** There is however a risk that a natural distribution would mean there are likely to be certain subjects where the majority of providers receive a bronze. This could create real incentives to close such courses if student demand drops off, and so would impact on diversity of provision and reduce student choice. This should be tested carefully during extended piloting to avoid such unintended consequences.

8. Interdisciplinarity

8.1 It is difficult to see how students would be able to identify the course they are interested in from the three broad groups (General and others in sciences, Humanities and liberal arts, and Combined and general studies) which it is proposed will apply to multi-subject programmes. Nor can we see how they could make meaningful comparisons between such subjects at different institutions. We conclude, **this approach will not provide students with useful information and may in fact discourage institutions from offering interdisciplinary programmes.** This would have a negative knock-on impact for student choice and stymy opportunities for innovation.

8.2 In addition, DfE's analysis of the proportion of subjects this would affect (2%) is based on providers who entered TEF in 2016/17 and so does not take account of many Scottish institutions where the provision of flexible modular degrees is more common.

9. Additional evidence

9.1 Given the significant additional burden which subject-level TEF assessments are expected to place on institutions – and by extension, students and taxpayers – **it would be helpful to explore whether PSRB accreditation for certain subjects could lead to the removal of the need for a separate TEF rating**, or at least a lighter-touch assessment process.

9.2 We would of course need to be cognisant that standards set by their accreditation schemes vary significantly so it may not make sense to place equal weight on all PSRB assessments. A way would need to be found of identifying those PSRBs where accreditation processes are sufficiently robust (and perhaps central enough to the content and delivery of the course) to replace the need for subject-level TEF assessment. It would also be important to ensure the use of PSRB accreditation as part of the TEF did not skew the relationship between universities and such bodies, leading to unnecessary micro-management.

10. Duration of award

10.1 Whilst extending the duration of an award (from five to six years) could reduce the administrative burden of the exercise, it could limit the usefulness of the information provided to students as a result of the historic nature of the metric data. Any extension in the minimum re-application period would mean providers whose performance on the metrics had improved in the interim would need to wait longer to have the opportunity to uprate their award. **It may be that a rolling process would be more appropriate with different subjects assessed in different years to spread the burden.**