

## Annex F: Proposed guidance to the TEF panel on assessing evidence

This annex relates to proposal 11.

1. This annex sets out further information – in addition to the issues covered under proposal 8 – that would be included in guidance to the TEF panel members on how to interpret and weigh up the evidence and form rating judgements. We have not prepared full draft guidance at this point as it is our intention to develop it in accordance with the decisions that follow from this consultation.
2. We welcome comments on the content of this annex as well as the proposals set out in the main consultation.
3. We propose to guide panel members to:
  - a. Initially, interpret the available evidence to **identify very high quality and outstanding quality features** within each aspect.
  - b. Then, consider a **rating for each aspect** by weighing up all the evidence relating to each aspect and applying the criteria for aspect ratings.
  - c. Finally, consider the **overall rating** for the provider, applying the criteria for the overall ratings.
4. We do not propose to specify the precise order in which evidence should be considered. We would anticipate that panel members would first familiarise themselves with the package of evidence for a provider and its context, and then refer iteratively to different pieces of evidence as they work through their assessment. It would be likely that they would consider the indicators relating to all a provider's students (in each mode of study) and the submissions at an early stage, before considering in detail the split indicators. In all cases they would make judgements having weighed up all the evidence.
5. Throughout the assessment, panel members would take account of a provider's context. This would be reflected in each source of evidence, including that:
  - a. The submissions could include narrative information about the provider's mission and context.
  - b. The TEF indicators would include data about the size and shape of provision. This could be supplemented by further data in the submissions.
  - c. The TEF indicators would show the provider's performance in relation to its benchmarks, which take account of a range of student and course characteristics.
  - d. Evidence in the provider's and students' submissions would be relevant to the mix of students and courses at the provider.

## Identifying very high quality and outstanding quality features

6. Having familiarised themselves with the provider's context, panel members would consider the available evidence to identify 'very high quality' or 'outstanding quality' features (see Annex B) and consider how far they apply across all the provider's student groups and the range of its courses and subjects. They would consider the evidence in both the submissions and the indicators, testing these sources of evidence against each other and weighing them up to identify 'very high quality' or 'outstanding quality' features.

### Interpretation of evidence in submissions

7. In considering the **provider and student submissions** panel members would draw on their expertise to interpret and weigh up whether the evidence suggests that a feature or features may be:
  - a. **Outstanding quality**, where there is sufficient evidence that the quality of the student experience or outcomes are among the very best in the sector, for the mix of students and courses concerned.
  - b. **Very high quality**, where there is sufficient evidence that the quality of the student experience or outcomes are materially above any relevant baseline quality requirements, for the mix of students and courses concerned.
8. In considering how compelling the evidence in a **provider submission** is, and how much weight to place on it, we propose the panel would consider the extent to which:
  - a. The evidence is directly relevant to a provider's mix of students and courses. Evidence would be more compelling, and greater weight placed on it, where the submission demonstrates the provider has a clear understanding of its students, tailors its approaches to its mix of students and courses, and demonstrates impact for its students.
  - b. Policies and practices are evidence-based, and their impacts are demonstrated. Evidence would be more compelling, and greater weight placed on it, where a provider's policies and practices are informed by robust evidence, data or evaluation, and there is robust evidence of the impact of these policies and practices in terms of delivering an excellent student experience or student outcomes.
  - c. The evidence overall covers all a provider's student groups and courses within the scope of the TEF assessment. Evidence relating to particular groups of students or courses could be important, for example to demonstrate improvement of specific subjects or the impact of interventions targeted at particular groups of students. However, greater weight would be placed on the totality of evidence relating to a feature, where it covers all the provider's groups of students and courses that are in scope of the TEF.
  - d. The evidence is relevant to the features of excellence related to that aspect. Greater weight would be placed on evidence that is directly relevant to these features, although to ensure the assessment can recognise diverse forms of higher education the proposed features are not intended to be exhaustive. The panel would also consider other evidence presented that is relevant to the quality of the student educational experience and student outcomes.

9. In considering how compelling the evidence in a **student submission** is, and how much weight to place on it, we propose the panel would consider the extent to which:
- The evidence reflects the views of students within the scope of the TEF assessment. Evidence would be more compelling, and greater weight placed on it, where it clearly articulates the views of students, and is broadly representative of all student groups and courses within the scope of the TEF assessment.
  - The evidence is relevant to the features of excellence related to that aspect. Greater weight would be placed on evidence that is directly relevant to these features, although to ensure the assessment can recognise diverse forms of higher education the proposed features are not intended to be exhaustive. The panel would also consider other evidence presented that is relevant to the quality of the student educational experience and student outcomes.

### Interpretation of indicators

10. When reviewing the **indicators**, panel members would be guided to interpret performance as indicative rather than determinative. We propose the panel would interpret a provider's indicators as initial evidence, to be tested against evidence in the submissions, of:
- Outstanding quality**, where the indicator is materially above the provider's benchmark. We consider this would indicate that a provider's performance may be among the highest quality in the higher education sector, for its mix of students and courses.
  - Very high quality**, where the indicator is broadly in line with the provider's benchmark. This is because we consider that the English higher education sector is generally high-performing, and many providers deliver an educational experience and support their students to achieve outcomes that are among the best in the world. This is in line with our views set out in our consultation on regulating student outcomes. A provider's performance that is broadly in line with its benchmark would generally indicate that its performance, for its mix of students and courses, was in line with that achieved by other providers in the sector. We propose that we should consider this to be recognised as very high quality.
  - Not very high quality**, where the indicator is materially below the provider's benchmark. We consider that this would not indicate very high quality for its mix of students and courses. However, this would not be taken as definitive evidence that the feature to which the indicator is relevant is not very high quality. The panel members would consider any relevant evidence or further context relating to the indicator within the submission before making a judgement.
11. To support consistent interpretation of the indicators, we propose to define 'materially above' and 'materially below' benchmark in a consistent way. We have considered what values would appropriately signify a material difference from a provider's benchmark and propose that:
- performance that is at least 2.5 percentage points above benchmark should be considered as materially above benchmark
  - performance that is at least 2.5 percentage points below benchmark should be considered as materially below benchmark

- performance that is within 2.5 percentage points of the benchmark in either direction should be considered as broadly in line with the benchmark.
12. In addition, we recognise that in some cases, a provider's benchmark may be so high that it would be difficult for the provider to materially exceed its benchmark. We propose that where a provider's benchmark for continuation is 95 per cent or higher, and the provider is not materially below its benchmark, the panel would interpret this initially as evidence of **outstanding quality**.
  13. Our supporting analysis and reasons for identifying these values are set out in 'Materiality and high benchmark values for use in interpretation of the TEF indicators'.<sup>73</sup>
  14. When interpreting the indicators, the panel would consider the level of statistical uncertainty in the position of the provider's indicator against its corresponding benchmark. We propose to use 'shaded bars' to represent the distribution of statistical uncertainty around the difference between the provider's indicator value and its benchmark. Guiding lines included on the display of the shaded bars would indicate where performance could be considered as materially above or below benchmark. Proposals for how this would be presented are described under proposal 9, and are demonstrated in the example 'TEF by provider' data dashboards that we have published alongside our consultation on indicators.<sup>74</sup>
  15. We propose to guide the panel to take account of statistical uncertainty by considering the position of the 'shaded bar' in relation to the 'guiding lines', recognising that the bar may cross one or both of these lines. We propose the panel should interpret the strength of the statistical evidence by using four indicative categories, aligned with those in proposal 6 in our consultation on regulating student outcomes. These categories are deliberately not discrete, as they describe the strength of statistical evidence, which is on a continuous scale, and are designed to avoid arbitrary divisions. The four categories are described with reference to statistical confidence<sup>75</sup> as follows:
    - a. Around 99 per cent statistical confidence would provide compelling statistical evidence.
    - b. Around 95 per cent or higher statistical confidence would provide very strong statistical evidence.
    - c. Around 90 per cent or higher statistical confidence would provide strong statistical evidence.
    - d. Around 80 per cent or higher statistical confidence would provide probable statistical evidence.
  16. These categories would be used when considering how far a shaded bar is materially above, broadly in line with, or materially below the benchmark. For example:

---

<sup>73</sup> See [www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/the-tef/](http://www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/the-tef/).

<sup>74</sup> Available at [www.officeforstudents.org.uk/data-and-analysis/student-outcomes-and-experiences-data-dashboards/tef-by-provider/](http://www.officeforstudents.org.uk/data-and-analysis/student-outcomes-and-experiences-data-dashboards/tef-by-provider/).

<sup>75</sup> In this context, statistical confidence should be interpreted from the supplementary table shown alongside the shaded bars. The table reports three figures, each one showing the proportion of the distribution of statistical uncertainty, represented by the shaded bar, that is materially above, broadly in line with, or materially below benchmark.

- a. If 90 per cent of the distribution represented by shaded bar is above the guiding line for 'materially above benchmark' this would provide strong statistical evidence that the provider's performance is materially above its benchmark. It would be interpreted as strong initial evidence of an outstanding feature.
  - b. If 95 per cent of the distribution represented shaded bar is between the two guiding lines, this would provide very strong statistical evidence that the provider's performance is broadly in line with its benchmark. It would be interpreted as very strong initial evidence of a very high quality feature.
17. While we consider the NSS indicators important we do not consider they provide direct measures of the proposed features of the student experience. The NSS indicators would therefore be interpreted as providing part of the evidence needed to identify very high quality or outstanding quality features of the student experience. They would need to be supplemented with further evidence of excellence in the submissions.
  18. We consider the outcome indicators provide more direct measures of positive student outcomes. We have proposed some very high quality and outstanding features within the student outcomes aspect (see Annex B) that could be identified without necessarily requiring further evidence in the submission. For example, continuation and completion rates that are materially above a provider's benchmarks could be sufficient evidence for the feature 'There are outstanding rates of continuation and completion for the provider's students and courses'. However, where these indicators are below a provider's benchmark, this would not be determinative that the associated feature is 'not very high quality'. The panel members would consider any relevant evidence or further context relating to this indicator within the submission before making a judgement. This partly recognises that the factors we propose to include in calculating benchmarks do not include all possible factors that could have affected a provider's historical performance.
  19. The panel members would primarily consider the indicators that cover all the provider's undergraduate students and courses, within each mode of study. This is because the TEF rating is intended to represent the overall quality of all the provider's courses and students in scope for the TEF assessment. The indicators for each mode of study represent overall performance across all types of courses, subjects and student groups, weighted according to the number of students in each category.
  20. Secondly, the panel members would consider the 'split' indicators within each mode of study. As proposed in the indicators consultation, the split indicators would be presented by student characteristics, subject, teaching arrangements, type and level of course, and year (see proposal 2 of the indicators consultation for details on the reporting structure, and proposal 9 for details relating to the types of split indicators). By considering the split indicators we intend for the incentives created by the TEF to apply across the range of student groups, course types and subjects at each provider. We propose that panel members would consider the split indicators in order to:
    - a. Consider how far very high quality and outstanding quality features might apply across all a provider's student groups and range of courses and subjects.

- b. Test the evidence in a provider's submission about its strengths and areas for improvement, including the provider's own analysis and use of the split indicators, alongside any other evidence it determines for itself.
21. The OfS's regulation of access and participation focuses on reducing gaps between student groups. As we propose that the TEF should reinforce but not duplicate our regulation of access and participation, we propose that when considering student characteristic splits the TEF panel members would not focus on gaps between student groups within a provider. Instead, TEF panel members would use the student characteristic splits to consider how far the provider delivers excellence for all its groups of students, relative to its benchmarks for each of those student groups. The panel members would also consider these splits to test evidence in the provider's and student's submission about equality of the student experience and outcomes.
22. We propose that panel members would consider the year splits in combination with evidence in the provider and student submissions, to test evidence about the impact of the pandemic in particular years, or evidence of improvements to the student experience or student outcomes, within the assessment period.
23. We propose that TEF ratings would reflect the quality of the student experience and student outcomes over the four-year period covered by the assessment as a whole, without weighting the years differentially. We have considered whether it would be appropriate to give greater weight to more recent years over earlier years within the four-year period, however we do not propose to do so. This is because we intend that the TEF should create incentives for improvement and excellence that apply continuously (before, during and after assessment exercises) rather than acting more strongly in some years and less strongly in others.
24. Where an indicator is not reportable (see proposal 9), or where there is a high degree of statistical uncertainty in the indicator, panel members would need to place proportionately greater weight on evidence in the submissions to identify a feature as very high quality or outstanding. The onus would be on a provider to ensure there is sufficient evidence of excellence in its submission.

## **Aspect ratings**

25. At this stage panel members will have identified very high quality and outstanding quality features in each aspect and considered how far they apply across the range of students and courses at the provider. Then, panel members would weigh up all the evidence relating to each aspect as a whole, and apply the criteria to make a judgement about the rating for each aspect. The proposed criteria for aspect ratings are in Annex B. We propose that in doing so:
- a. Overall, the indicators would contribute no more than half of the evidence of very high quality or outstanding features, for the aspect. This is because we consider that the indicators based on NSS are important but are not direct measures of the features covered by the student experience aspect; and the student outcome indicators provide measures for only some of the features of the student outcomes aspect.
  - b. Panel members would consider the extent to which the very high quality and outstanding quality features are evident across all groups of students, subjects and course types at the provider.

- c. Panel members would consider the extent to which there are very high quality and outstanding quality features across the aspect as a whole, rather than treating the features as a checklist. Beyond the indicators, a provider could choose to place more or less emphasis on particular features depending on their relevance to its context. For this reason, we propose that the panel members would not assign any predetermined weight to, or equally weight, each feature. In order to consider awarding the highest rating the panel would not, for example, require there to be equally strong evidence across all features, so long as it judges there is evidence of typically outstanding quality across the aspect as a whole.

26. Having considered these issues and weighed up all the evidence relevant to each aspect, panel members would be guided to make a 'best fit' judgement against the criteria for the aspect ratings. The proposed criteria for the aspect ratings (as set out in Annex B) do not seek to describe how every possible combination of very high quality and outstanding quality features would be associated with a particular rating category. It is intended that the panel would use the ratings criteria to decide which of the rating categories is a 'best fit', meaning that the criteria for that rating category are, on the whole, more applicable to all the evidence than the criteria for any other rating category.

### **Absence of excellence**

- 27. The panel would not award a rating to an aspect where it judges there is an absence of very high quality or outstanding features across that aspect. This would be the case where there are no or minimal very high quality or outstanding features found in that aspect.
- 28. The panel might also find features of the student experience or outcomes that it considers to be clearly below the level of 'very high quality', or that may be of concern. This may be alongside other features at a provider that it considers to be very high quality or outstanding. In such cases the panel would consider if those features that are clearly below the level of 'very high quality', or that may be of concern, are sufficiently serious or widespread to prevent it from concluding that overall, there is sufficient evidence of excellence to award a rating of Bronze or above.
- 29. The following non-exhaustive list provides examples that could be considered as clearly below the level of 'very high quality' or that may be of concern:
  - a. If a provider's continuation or completion rates are materially below its benchmarks, and the information in the provider's submission does not adequately explain why this is the case or set out an effective approach the provider has in place to support its students to succeed in their studies.
  - b. If there are split indicators that are materially below benchmark for some groups of students and materially above benchmark for others, and the submission does not adequately explain why this is the case or set out an appropriate approach the provider has in place to deliver very high quality courses for all its groups of students.
  - c. If there are split indicators that are materially below benchmark for some subjects and materially above benchmark for others, and the information in the submission does not adequately explain why this is the case or set out an effective approach the provider has in place to deliver very high quality courses in those subjects that are below benchmark.

- d. If the provider's submission does not adequately explain how it engages with its students to ensure a very high quality experience; and the student submission provides reasonable evidence that the provider does not meaningfully do so.
- e. If the submission does not adequately articulate what educational gains the provider intends for its students, or how it supports its students to achieve them.

## Overall ratings

- 30. We propose that the rating for each aspect would be considered before determining the overall rating for the provider, and that the process for determining the overall rating would rely on a combination of rules and expert judgement.
- 31. Where each aspect is awarded the same rating, we propose that the overall rating would also be the same (or if neither aspect is awarded a rating, there would be no overall rating). This is because where the same rating is awarded to both aspects, it follows that overall, the student experience and student outcomes would be at the same level.
- 32. Where each aspect is awarded a different rating, we propose the following two rules should apply:
  - a. The overall rating should not be higher than the highest aspect rating.
  - b. The overall rating should be no more than one rating higher than the lowest aspect rating.
- 33. We propose these rules to ensure that the overall rating is coherent with the aspect ratings, and to support consistent decision-making. We have also had regard to the government's guidance that student outcomes should act as 'limiting factors', meaning that a provider should not be able to achieve a high TEF rating if it has poor student outcomes. We consider that the proposed rules would ensure appropriate limits to the overall rating, based on both the quality of the student experience and student outcomes.
- 34. Within these rules the panel would exercise its expert judgement. For example, if a provider had aspects rated Gold and Silver, the rules would not determine whether the overall rating would be Gold or Silver. It would be for the panel to make an overall 'best fit' decision. We propose the panel would do this by:
  - a. Equally weighting the two aspects.
  - b. Considering all the evidence across all features, and across all of the provider's student groups, subjects and courses, to make a 'best fit' judgement against the ratings criteria (see Annex B).
- 35. The effect of these proposals is set out in Figure 5. It shows that one effect of these proposals is that a provider that is awarded a rating for one aspect but not the other could still be considered for a Bronze overall rating.



Figure 5: Relationship between aspect and overall ratings

Student experience aspect rating	Student outcomes aspect rating			
	No rating	Bronze	Silver	Gold
	Overall rating			
	No rating	No rating or Bronze		
	Bronze	Bronze	Bronze or Silver	
Silver	No rating or Bronze	Bronze or Silver	Silver	Silver or Gold
Gold			Silver or Gold	Gold

If you have feedback on this annex, please respond to question 10.