



Department  
for Education

# **Teaching Excellence and Student Outcomes Framework: Subject- Level**

**Technical document to support the  
government consultation**

# Contents

1	About the consultation	3
2	Executive Summary	5
3	What is TEF?	6
4	Subject classification system	7
5	Duration of award	11
6	Overview of subject-level TEF design	14
6.1	Introduction	14
6.2	The two models	15
6.3	Model A: By exception	16
6.4	Model B: Bottom-up	19
6.5	Comparing the models	22
7	Model A: Generating exceptions	24
8	Model A: Relationship between provider and subject assessment	27
9	Model B: Relationship between provider and subject assessment	29
10	Metrics	32
10.1	Introduction	33
10.2	Grade inflation	36
10.3	Distribution of subject ratings	36
10.4	Non-reportable metrics	45
11	Additional evidence	57
12	Interdisciplinarity	62
13	Teaching intensity	66
13.1	Introduction	66
13.2	Rationale for measuring teaching intensity	67
13.3	Measures of teaching intensity	70
14	Next steps and relationship with independent review	80
	Annex A: Subject groupings	82
	Annex B: Absolute value thresholds	84
	Annex C: Metrics charts	86
	Annex D: Student research outline	92
	Annex E: GTQ example	95

# 1 About the consultation

**This is a technical document that is intended to be read alongside the Government consultation on [‘Teaching Excellence and Student Outcomes Framework: Subject-Level’](#)**

This paper provides further detail and analysis about the consultation and the key proposals being made for subject-level TEF. The paper is structured as follows:

- The chapter headings and titles are aligned to those in the consultation document.
- Chapters 1 to 3 are introductory and as such, there is generally limited additional technical detail to be added.
- Chapters 4 to 13 provide further detail on each of the topics in the consultation document, with the chapter heading numbers for each topic aligned to those in the consultation document. Each chapter in this paper provides detailed information and analysis on the background, proposals and rationale for each of the consultation topics.

## Student research

- 1.1 We have commissioned some research with students to test what the best subject classification system would be and to seek views on what is important to students when choosing an institution.
- 1.2 See ‘Annex D: Student research outline’ for more information.

## Focus of the consultation

- 1.3 We held a series of collaborative design workshops with sector bodies to develop the initial design of subject-level TEF. These discussions covered topics such as the rationale for moving to subject-level TEF, how the TEF framework applies at subject-level and what models could be piloted. These discussions helped to inform the overall design of subject-level TEF and we are very grateful to all those who contributed to those sessions.
- 1.4 We now intend to move beyond these higher level principles to focus on the detailed design of subject-level TEF and key elements of the two models being proposed.

- 1.5 **The consultation therefore focuses on particular topics that we would like to explore in detail and poses specific questions on which we are seeking feedback.**
- 1.6 **All of the topics covered in the consultation focus specifically on the design of subject-level TEF.** The consultation is not intended to cover topics or issues related to provider-level TEF or elements of the TEF Specification. A technical consultation was undertaken on provider-level TEF during development of the policy for 'TEF Year Two' (assessments in academic year 2016/17). A lessons learned exercise was also undertaken on provider-level TEF following the release of the Year Two outcomes. Further consideration of both provider and subject-level TEF will occur as part of the independent review.

### **Who the consultation is for**

- 1.7 The consultation seeks feedback from:
- Higher education providers
  - Representative bodies of higher education providers
  - Professional, Statutory and Regulatory Bodies (PSRBs) and other subject specific bodies
  - Current and prospective students of higher education
  - Employers
  - Teachers, careers advisers and parents
  - Other stakeholders with an interest in higher education policy.
- 1.8 The consultation is open to all in the United Kingdom and we are seeking the views of the devolved administrations.

### **Additional copies**

- 1.9 Additional copies are available electronically and can be downloaded from [GOV.UK DfE consultations](https://www.gov.uk/dfe-consultations).

## **2 Executive Summary**

2.1 There is no further technical detail for this chapter.

### **3 What is TEF?**

- 3.1 There is no further technical detail for this chapter. For more information about TEF please refer to the [TEF Specification](#).

## 4 Subject classification system

We will use a **subject classification system** to define what, for the purposes of assessment, a 'subject' is. We propose to use the second level of the Common Aggregation Hierarchy (CAH2) to define 'subjects' for the purpose of subject-level TEF. The CAH2 is made up of 35 subjects. We expect that this classification of subjects will be meaningful for students as it has been designed around teaching and with student information in mind. In Model B, we will group the 35 subjects into 7 subject groups to reduce the burden of assessment.

### Background

- 4.1 In subject-level TEF, we need to define the level at which assessment and ratings will take place. We propose using a subject classification system to define what, for the purposes of assessment, a 'subject' is.
- 4.2 The diversity of the sector makes choosing a subject classification system difficult. As autonomous institutions, providers choose their own subject classifications to define, for example, their courses, programmes and their organisational structure. The chosen subject classifications vary significantly across institutions. As such, we recognise that no single system will be perfect for every institution.
- 4.3 In considering the level of granularity at which subjects should be assessed and given ratings, we have considered at which level of granularity subjects will:
  - be clearly understood by students making choices about what to study
  - group courses that are likely to be reasonably similar in teaching quality (although there will always be courses that straddle subject boundaries)
  - have large enough numbers of students that metrics are reportable
  - not be so granular that the time and other resources required from providers and panellists becomes unmanageable.
- 4.4 Some stakeholders have asserted that students would ideally like information at course level. Whilst students clearly need some information at course level (for example, on the Unistats website), TEF assessment at this level would be impractical. Too much aggregation, however, will also be meaningless for students (such as presenting science as one subject). The subject classification system therefore needs to strike a balance between the two. Ideally, it should also be based on an existing system of subject classification, to avoid proliferation of systems and consequent confusion.

- 4.5 Our analysis has confirmed that the metrics are likely to continue to be reportable for the vast majority of students if they are reported for up to approximately 40 subjects. The proposal and options outlined below are therefore based on a subject granularity of up to approximately 40 subjects.

## **Proposal and rationale**

### **Common Aggregation Hierarchy Level 2**

- 4.6 When we spoke with sector representatives, there was a strong consensus that it would be greatly preferable for the TEF to use an existing subject classification system rather than to create a new one. We would also like to use, and align the TEF, with a system that is widely used and understood by the higher education sector.
- 4.7 We propose using the second level of the Common Aggregation Hierarchy (CAH2) to define 'subjects' for the purpose of subject-level TEF. This would apply to both Model A and B (see Chapter 1 for a description of the two models). The 35 subjects that make up CAH2 can be found in 'Annex A: Subject groupings'.
- 4.8 The Higher Education Statistics Agency (HESA) developed the CAH as an aggregation system to sit alongside the Higher Education Classification of Subjects (HECoS). The CAH and HECoS will together replace the Joint Academic Coding System (JACS), which is currently used by the Universities and Colleges Admissions Service (UCAS) for students applying to university as well as on the Unistats website and by HESA when reporting data. CAH also aligns with the system of subject benchmarking currently used in the TEF. We expect that this will be meaningful for students, as it has been designed around teaching and with student information purposes in mind. The student research we are conducting will test this with prospective students. HECoS will be implemented for the academic year starting in the autumn of 2019. We currently expect full implementation of subject-level TEF in academic year 2019-20. This means providers will already be using CAH2 by the first year of subject-level TEF assessments.
- 4.9 We worked with HESA on the development of the CAH in order to make it suitable for the purpose of the TEF. Level 2 has 35 subjects and HESA have taken into account the number of students in each subject (which is important for metrics purposes), clarity for students and alignment with the Research Excellence Framework (REF) where there is no strong reason to deviate from it.
- 4.10 Our analysis, using the TEF data from assessments in academic year 2016/17 ('TEF Year Two'), has confirmed that, using the 35 subjects in CAH2, there are approximately 4,500 instances of subjects being offered by providers across the

sector as a whole<sup>1</sup>. Subject instances are identified by looking at the number of subjects offered by each provider and adding this up across the sector. For example, if provider 1 offers 3 subjects and provider 2 offers 5 subjects, then there are 8 subject instances in total. It is useful to use subject instances as it allows us to determine the number of assessments that could take place. Across these 4,500 instances, 97.7% of students are included in subjects with reportable metrics, suggesting that subject-level TEF results would have meaning for nearly all applicants.

## Alternatives

4.11 We also considered the Units of Assessment used in the Research Excellence Framework (REF UoAs) to classify subjects. This was discounted because we felt it was unsuitable for students as its groupings are heavily tailored towards research. Using the REF UoAs for the TEF would send a contradictory message about our commitment to increasing the parity of esteem between research and teaching, because it would mean research considerations would be driving the TEF as well as the REF.

## 7 subject groups

4.12 As discussed in Chapter 6.4, in Model B we propose to group the 35 CAH2 subjects into 7 groups for the purpose of submissions (see Annex A: Subject groupings). This should streamline the assessment process and reduce the burden for both panel members and providers. Individual subjects will still receive individual ratings: the groups are administrative constructs for the purpose of assessment and will not be used in communicating ratings to students.

4.13 Given the diversity of the sector, we recognise that these 7 subject groups will not suit every institution's structure. When grouping the subjects we considered what we thought would be most representative of the sector as a whole, acknowledging that this will not reflect the specific design of every institution.

4.14 These groups have also been designed to group together subjects that we believe are likely to have similar teaching practices, teaching quality, and student outcomes.

4.15 Each subject group is treated the same in the assessment process. Therefore, we consider it appropriate that each group is fairly even in terms of their relative size and breadth. The table below provides some information on this using

---

<sup>1</sup> The number of providers that were in scope for TEF during the TEF assessments that took place in academic year 2016/17 was used to calculate the number of subject instances. It should be noted that this includes data from providers who were eligible for full TEF awards, and differs from the number of providers who received a TEF award in that assessment year.

contextual data from TEF assessments taking place in academic year 2017/18. In general, this shows that the 7 subject groups are relatively well balanced in terms of size.

4.16 Table 1 shows the number of providers and students that were represented in each of the 7 broad subject groups. While there was some variation across the groups in this table, no subject group represented a significantly larger proportion of students than any other. There were 3 subjects each with around 20% of the student population and 4 subjects each with around 10% of the student population. We also see variation in the number of providers that offer courses in each of the 7 subject groups. For example, 257 providers offered courses under the Natural sciences subject group, while 434 providers offered courses under the Social sciences group (Table 1).

**Table 1: The 7 subject groups by number of providers and number of students**

<b>Subjects</b>	<b>Number of providers</b>	<b>Number of students</b>
Arts	379	163,865 (9%)
Business and law	420	336,270 (19%)
Engineering and technology	397	218,870 (12%)
Humanities	296	242,140 (13%)
Medical and health sciences	378	369,880 (20%)
Natural sciences	257	168,385 (9%)
Social sciences	434	311,880 (17%)

**Note:** Sourced from the TEF contextual data prepared for assessments in academic year 2017/18 (TEF Year 3). Students whose course subjects span more than one subject, their contribution is split across the two according to the broad balance of subjects ie a Maths and Modern Dance student would typically count as a half in each of natural sciences and arts. There was 1,811,290 students in total in the TEF contextual data. The TEF contextual data includes the aggregation of 3 successive years of student data reflecting the full undergraduate student cohort from 2013/14 to 2015/16.

4.17 We will be testing these groupings in the pilots. Providers will have the opportunity to move one subject in and one subject out of each group in order to reflect better the make-up of their specific structures.

## 5 Duration of award

We are proposing to extend the **duration and re-application period** of awards under subject-level TEF to reduce the administrative and financial burden on providers and assessors. We are consulting on two options for the length of the duration and re-application period:

- A 5 year duration with a 3 year minimum re-application period, with an annual assessment process.
- Running the assessment process every second year and in combination with this, a duration of 6 years and re-application period of 4 years.

### Background

- 5.1 This section considers whether we should change the duration and re-application period of awards under subject-level TEF. Given that subject-level TEF includes both provider and subject-level ratings, and there is a relationship between these ratings in both design models (see chapter 1), the duration and re-application periods will need to be consistent across both provider and subject ratings.
- 5.2 The current provider-level TEF award is valid for a period of 3 years, subject to a provider continuing to meet eligibility requirements, unless a provider does not have the requisite 3 years of core metrics to inform the assessment. For a provider that has only 1 or 2 years of core metrics, the award granted lasts for 1 or 2 years respectively. A provisional TEF award given to a provider that does not have suitable metrics lasts for 1 year. The assessment process runs every year and providers are able to re-apply each year if they wish to do so.

### Proposal and rationale

- 5.3 **We propose to extend the award duration and the re-application period.**
- 5.4 There are **two options** for extending the award duration and re-application period for subject-level TEF:
1. Annual assessment cycle, with duration of 5 years and re-application every 3 years
  2. Assessment every two years, with duration of 6 years and re-application every 4 years.
- 5.5 This refers to the award duration for providers with three years of core metrics at provider-level. As is currently the case, we anticipate that for providers that have

only one or two years of core metrics at provider-level, the award duration would be shorter.

## Option 1

5.6 In the 'TEF Lessons Learned from Year Two' publication we anticipated that TEF would move to a 5-year cycle. In option 1, **TEF ratings would be awarded for 5 years** for a provider with the requisite data, with a **minimum re-application period of 3 years**. This period aligns with the 3 years of aggregated data used in TEF assessments.

5.7 In this option, **the assessment process would be run annually**.

## Option 2

5.8 The pilots are testing how an annual subject-level assessment will work in practice. Based on early feedback from the Higher Education Funding Council for England (HEFCE), it is possible that the assessment process for subject-level TEF may take slightly longer than a year. Given the two stage process (i.e. provider- and subject-level assessments), additional time may be needed to undertake moderation and verification, providers may need longer to produce submissions and the time to draft statement of findings and collate the outcomes may be extended.

5.9 Given these considerations, an alternative would be to **run the assessment process every two years**, with the duration and re-application period aligned to this. If TEF assessments were run every second year, we could have:

- An award duration of 6 years
- A re-application period of 4 years.

## Rationale

5.10 Our **rationale** for extending the duration and re-application period in subject-level TEF is to account for the higher volume of work involved, and the administrative and financial burden that would be placed on both providers and assessors. Extending the duration and the re-application period will ease the burden of the application and assessment process so that the costs remain proportionate and the exercise continues to offer value for money both for the sector and for taxpayers. Preventing re-application every year will have a similar impact on cost, and responds to concerns about potential game-playing, in which a provider re-applies every year to try to get a higher rating, placing more effort on improving its submission than in making genuine improvements for students.

- 5.11 A potential **disadvantage** of extending the duration of the award is the impact on student information, in terms of how relevant and up to date TEF awards are. We recognise that an award given 4 or 5 years ago would be based on data that is not as recent. Given this, the award may not be fully representative of the current student experience.
- 5.12 These are very serious considerations, which we have balanced against the benefits of extending the award in reducing the burden and bureaucracy of the assessment process. On balance, we believe that **either option will deliver value for money for the sector and will be a proportionate approach to use for subject-level TEF.**
- 5.13 We have no view on which option is best. The pilots will help inform us as to which may work better in practice but we are also seeking wider sector views.

## 6 Overview of subject-level TEF design

**Two alternative models** are being proposed for subject-level TEF:

- **Model A:** A 'by exception' model giving a provider-level rating and giving subjects the same rating as the provider where metrics performance is similar, with fuller assessment (and potentially different ratings) where metrics performance differs.
- **Model B:** A 'bottom-up' model fully assessing each subject to give subject-level ratings, feeding into the provider-level assessment and rating. Subjects are grouped for submissions, but ratings are still awarded at subject-level.

### 6.1 Introduction

6.1.1 **The design of subject-level TEF is based on the provider-level framework.**

6.1.2 The lessons learned exercise following academic year 2016/17 of the TEF has demonstrated that the process has operated in a fair and transparent way. Evidence from the exercise indicates that assessors' findings were clear and robust, producing credible results that reflect teaching excellence across the sector.

6.1.3 Given this, we have retained the existing key elements of TEF such as; the criteria, the use of benchmarked metrics, provider submissions, the independent panel assessment process and the rating system.

6.1.4 Assessment will be holistic, based on both core and split metrics supplemented by additional evidence. It will be carried out by peers comprised of experts in teaching and learning as well as subject experts, student representatives, employer representatives and widening participation experts.

## Provider-level versus subject-level TEF

Under subject-level TEF, we would retain the provider-level assessment and rating. Therefore, participating providers would receive **both**:

- a provider rating, and
- a rating for each subject they teach (subject-level rating).

Once fully implemented, subject-level TEF would entirely replace the existing provider-level TEF process. We intend, at this point, for existing provider-level awards to continue to be valid for the remainder of their duration until the provider applies to subject-level TEF (subject to a provider continuing to meet eligibility requirements). However, TEF will be subject to a statutory independent review and we cannot prejudge the recommendations of the independent reviewer, whom we expect to report in Academic Year 2018/19.

- 6.1.5 To classify 'subjects' for assessment, we are proposing to use the 35 subjects set out in level 2 of the HESA Common Aggregation Hierarchy (CAH2). Participating providers would be required to submit their full provision for assessment. For providers offering all 35 subjects, this means they would receive 35 subject ratings, plus a provider rating.
- 6.1.6 To inform the assessment, subject-level TEF takes account of both metrics and submissions at both provider- and subject-level.

## 6.2 The two models

### 6.2.1 There are two alternative models being proposed for subject-level TEF:

- **Model A:** A 'by exception' model giving a provider-level rating and giving subjects the same rating as the provider where metrics performance is similar, with fuller assessment (and potentially different ratings) where metrics performance differs.
- **Model B:** A 'bottom-up' model fully assessing each subject to give subject-level ratings, feeding into the provider-level assessment and rating. Subjects are grouped for submissions, but ratings are still awarded at subject-level.

- 6.2.2 It is expected that, following feedback from the consultation and the first year of pilots, **one of these two models will emerge as the preferred option and will be taken forward in a refined form for the second year of pilots.** The Government does not currently have a preference between the two proposed models. The following two sections provide further detail on how each model

would work, and we are seeking views on which model you prefer and your rationale for that choice.

## 6.3 Model A: By exception

- 6.3.1 Model A is a 'by exception' model that gives subjects the same rating as the provider where metrics are similar to the provider's, with fuller assessment by exception. The model aims to:
- Reduce the burden that subject-level TEF could create
  - Focus fuller assessment on subjects that the metrics indicate may be performing differently and that may receive a different rating to the rest of the provider.
- 6.3.2 In Model A it is assumed that the rating produced at provider-level is indicative of the teaching quality and student outcomes for most subject areas. The metrics at subject-level will be used to indicate when this is not an accurate assumption and that a subject should be assessed, potentially resulting in a different rating. Model A triages subjects on the basis of metrics, recognising that for most subjects, the rating will be the same as the provider rating. On this basis, full assessment of every subject is not necessary and should be the exception, not the rule.
- 6.3.3 **All subjects whose metrics would cause the initial hypothesis (IH) for that subject to be different from the provider-level initial hypothesis (IH) would be treated as exceptions.**
- 6.3.4 Additionally, **providers would be able to choose a small number of non-exception subjects to be assessed** where they feel that they deserve greater scrutiny.
- 6.3.5 As in provider-level TEF, provider submissions in Model A will be limited to 15 pages. Subject submissions will only be produced for 'exceptions' and will be limited to 5 pages. All 10 criteria would apply in both provider and subject submissions. In responding to the consultation, while stakeholders are welcome to comment on the page limits for submissions, we believe the pilots are the best way to test whether these are appropriate.

6.3.6 Model A is shown visually in two diagrams on the following page. Figure 1 shows the conceptual design of the model and Figure 2 explains the process. As shown in these diagrams, the provider-level assessment and rating is determined first, using the same process as current provider-level TEF. Subject-level ratings are then determined in two different ways:

- **Non-exception subjects:** these subjects automatically receive the same rating as the provider-level rating (because the initial hypotheses for these subjects are the same as for the provider)
- **Exception subjects:** each of these subjects is assessed separately, through a holistic assessment based on subject-level metrics and a subject-level submission (because the initial hypotheses for these subjects are different from the provider).

Figure 1: Model A design

### 1. Provider-level assessment



### 2. Subject-level assessment

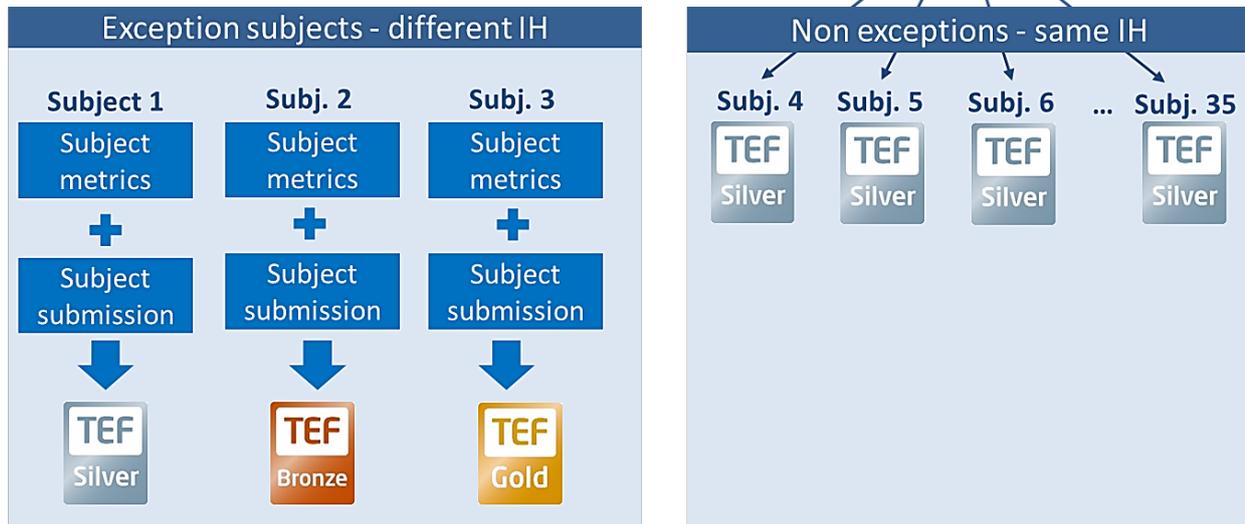
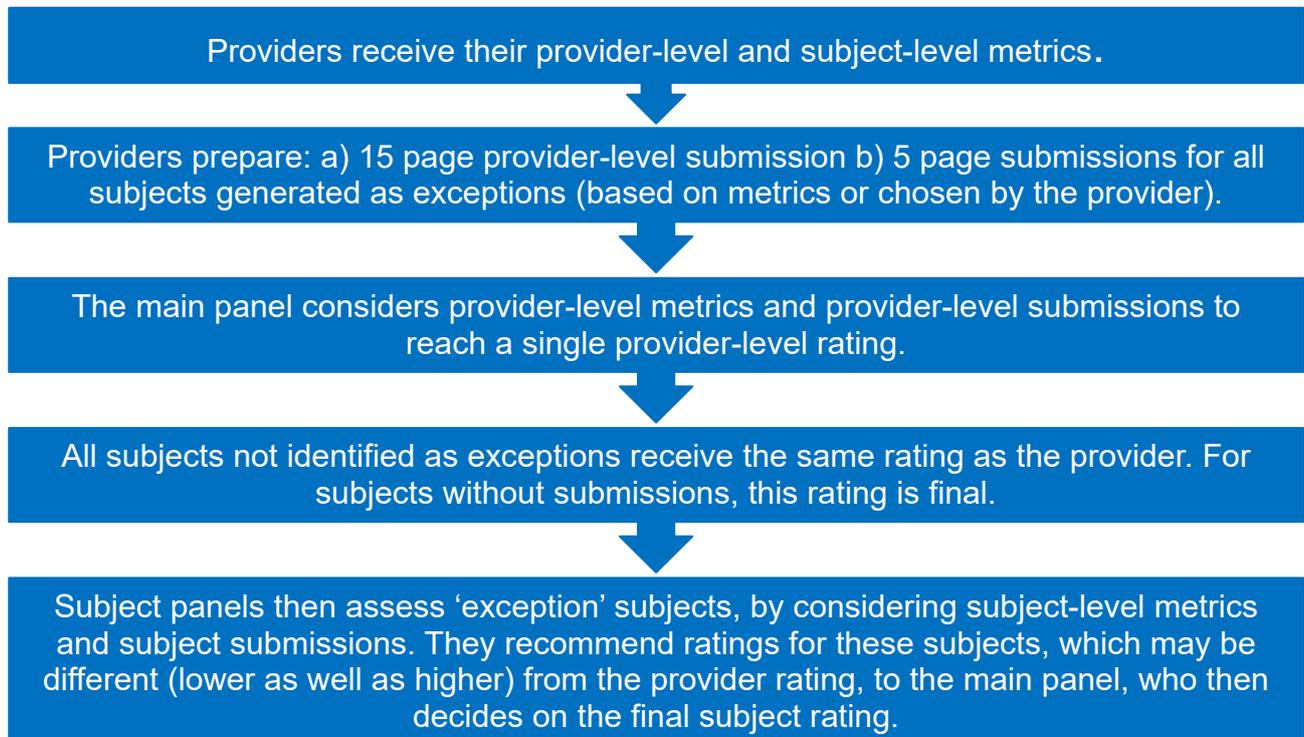


Figure 2: Model A process



## Model A: Single subject providers

- 6.3.7 In Model A, providers whose eligible provision falls entirely within one of the 35 CAH2 subjects will not have any exceptions generated, as the subject-level metrics will be identical to the provider-level metrics. Single subject providers will therefore produce only a single 15-page submission, which will be assessed by the provider-level main panel to give the same rating for provider and subject.
- 6.3.8 This page limit should not disadvantage single subject providers in terms of the space afforded to discussion of their provision, as the case for excellence at provider-level and subject-level will be aligned.
- 6.3.9 In responding to the consultation, while stakeholders are welcome to comment on the page limits, we believe the pilots are the best way to test whether these are appropriate.
- 6.3.10 We considered whether single subject providers would be disadvantaged if their subject was not assessed by a subject panel but went straight to the main panel. This would be the case if other providers had all their subjects assessed by subject panels, but in the 'by exception' model the majority of subjects will not be assessed by subject panels so we do not consider this to be a disadvantage.

## 6.4 Model B: Bottom-up

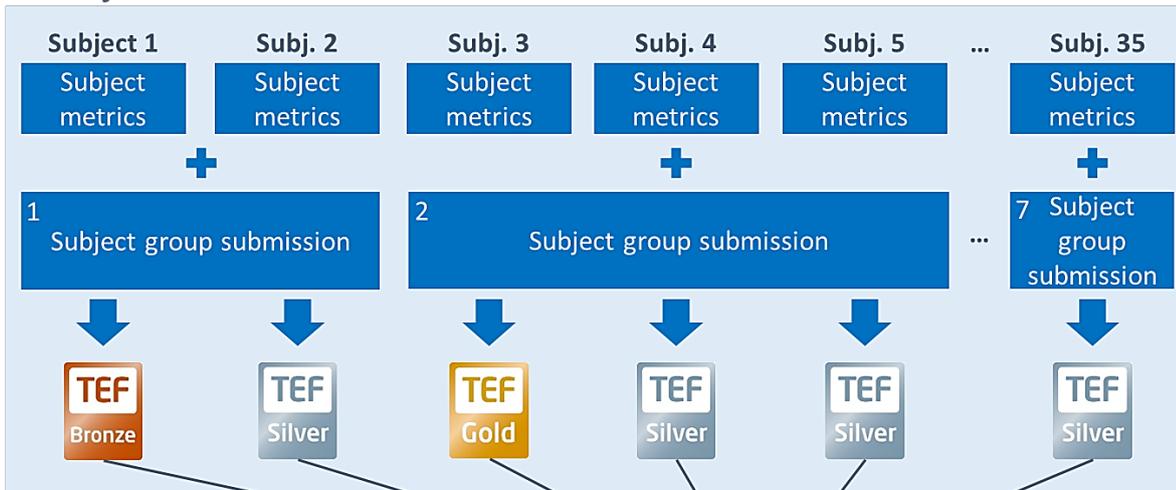
- 6.4.1 Model B is a 'bottom-up' model that assesses each subject fully, through metrics and submissions, and gives ratings for all subjects. These subject ratings are then used alongside provider-level metrics and a provider-level submission to determine the provider-level rating.
- 6.4.2 This model aims to:
- Demonstrate how provider-level ratings could be based in part on subject-level ratings.
  - Assess each subject fully in as manageable a way as possible, though we do recognise that full assessment of each subject will be a significant undertaking for providers.
- 6.4.3 **In Model B, all subjects are assessed.** For the purposes of submission writing, the 35 CAH2 subjects are aggregated into 7 subject groups. The 7 groups have been designed to group together subjects that are likely to have similar teaching practices, teaching quality, and student outcomes. This should streamline the process for both panel members and providers. For panel members, this means they will only need to review up to 7 submissions per provider, rather than up to 35. For providers, grouped submissions should help providers to streamline their

evidence and avoid duplication, and reduce the burden in areas where there is consistency between or across subjects.

- 6.4.4 We recognise that these groups may not fit each institution's specific make-up. Providers will therefore have the opportunity to move one subject in and one subject out of each group in order to better reflect the make-up of their specific structures.
- 6.4.5 Provider-level submissions will be limited to 10 pages. To account for the fact that providers will have different numbers of subjects in each group, group submissions will have variable maximum page length: 5 pages for each group in which a provider has 1 subject, with the page limit increasing by 1 page for each additional subject the provider has in that group. In responding to the consultation, while stakeholders are welcome to comment on the page limits for submissions, we believe the pilots are the best way to test whether these are appropriate.
- 6.4.6 All 10 current TEF criteria will be considered in subject-level assessments and should therefore be reflected in the subject-level ratings. As the subject ratings form part of the provider-level assessment process, the provider-level submission will be more limited than in Model A or provider-level TEF. It is proposed that the provider-level submission be focused purely on three criteria that are most relevant at an institutional level. These are:
- TQ2: Valuing Teaching
  - LE1: Resources
  - SO3: Positive Outcomes for All.
- 6.4.7 Model B is shown visually in two diagrams on the following page. Figure 3 shows the conceptual design of the model and Figure 4 explains the process. As shown in these diagrams, the subject-level assessment and rating is determined first. Subject-level ratings are determined by a holistic assessment of:
- Subject-level metrics
  - Subject-level submissions (written for 7 subject groups covering all 10 criteria).
- 6.4.8 The provider-level rating is then determined by a holistic assessment of:
- **Provider-level metrics** – These metrics are the same as provider-level TEF, with an initial hypothesis based on the core metrics
  - **Provider-level submission** – A 10 page submission focusing on 3 criteria
  - **Subject-level ratings** – The final subject-level ratings are used to form a subject-based initial hypothesis (see paragraph 9.5 for an explanation of this).

**Figure 3: Model B design**

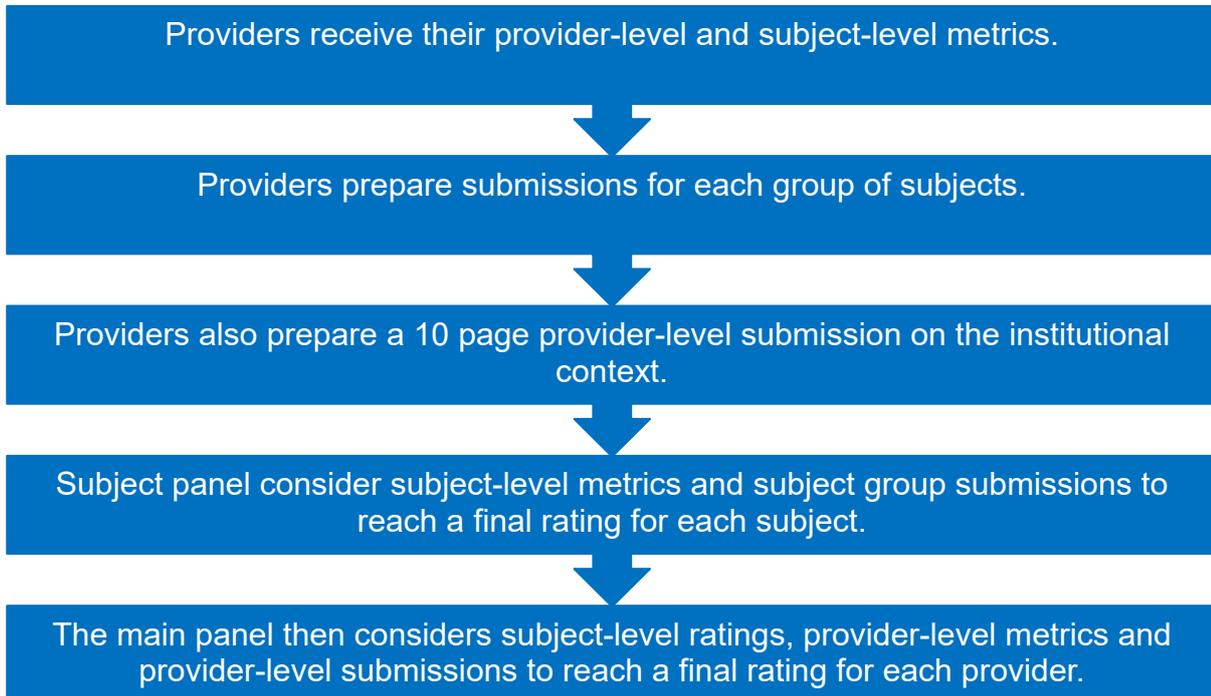
**1. Subject-level assessment**



**2. Provider-level assessment**



**Figure 4: Model B process**



## Model B: single subject providers

- 6.4.9 In Model B, single subject providers will produce a single submission with a maximum of 15 pages that will be assessed by the relevant subject panel, which will assign a rating. This rating will be the provider's subject-level rating and provider-level rating. However, provider-level assessors and the main panel will review these ratings fully, whereas their review of other subject-level ratings will be limited to moderation only.
- 6.4.10 While stakeholders are welcome to comment on page limits in responding to the consultation, we believe the pilots are the best way to test whether these limits are appropriate.
- 6.4.11 We will ask provider-level assessors and the main panel to identify any elements of a single subject provider's submission that are more akin to what other providers would include in their provider-level submissions. We will also ask them to consider whether the fact that this has been assessed at subject-level only would have had an impact on the rating given. If so, the rating would be adjusted accordingly.

## 6.5 Comparing the models

- 6.5.1 Table 2 demonstrates the similarities and differences between the two models. Both models use the CAH2 to classify subjects, and a provider-level rating is retained in both. The major difference is how subject and provider ratings are reached, by exception in Model A and from the bottom up in Model B.

**Table 2: Comparison of Model A and B**

	<b>Model A</b>	<b>Model B</b>
<b>Metrics</b>	Provider-level metrics will be released as per the TEF provider-level specification. Subject-metrics will be released for the 35 subjects and are the same for both models. The calculation and presentation of metrics and benchmarks will mirror the provider-level specification. Subjects will be identified using the second level of the Common Aggregation Hierarchy (CAH2). For single subject providers, subject-level metrics are identical to their provider-level metrics.	

	<b>Model A</b>	<b>Model B</b>
<b>Subject-level assessment</b>	Assessment at subject level is 'by exception'. Subjects are awarded the same rating as the provider, where the initial hypothesis is the same. There is fuller assessment (and potentially different ratings) for subjects where their initial hypothesis differs from the provider. Providers can choose a small number of additional subjects to be assessed.	Assessment is 'bottom-up': each subject is fully assessed to give subject-level ratings. These feed into the provider-level assessment and rating.
<b>Provider submissions</b>	A provider-level assessment is retained in both models	
	A provider-level submission of up to 15 pages (covering all 10 TEF criteria).	A provider-level submission of up to 10 pages, covering TEF criteria: <ul style="list-style-type: none"> <li>• TQ2: Valuing Teaching</li> <li>• LEI: Resources, and</li> <li>• SO3: Positive Outcomes for All</li> </ul>
<b>Subject submissions</b>	Subject submissions can include up to 5 pages per subject (where identified as exceptions by the metrics, or chosen as additional subjects).	Model B maps the 35 CAH2 subjects into 7 subject groups, which are used for submissions. These will have a variable maximum page length (between 5 and 13 pages) depending on the number of subjects the provider delivers within the subject group.
<b>Subject ratings</b>	Ratings are awarded for each of the 35 CAH2 subjects.	

## 7 Model A: Generating exceptions

In Model A we are proposing to generate exceptions by comparing the initial hypothesis in provider- and subject-level metrics. We also propose that providers would be able to select a small number of additional subjects to be assessed, allowing providers to put forward subjects that they believe deserve greater scrutiny.

### Background

- 7.1 In Model A it is assumed that the rating produced at provider-level is indicative of the teaching quality and student outcomes in most parts of the provider (i.e. in most subject areas). There will be some subjects however, where this assumption does not apply and the teaching quality and student outcomes are different. These subjects are deemed to be 'exceptions' to the underlying assumption that the provider-level rating is indicative of the quality at the subject-level.
- 7.2 Under Model A, full assessment of every subject is unnecessary, only those subjects identified as exceptions are assessed. The design of Model A therefore needs to include a method for identifying which subjects are sufficiently different to the provider to be treated as exceptions.

### Proposal and rationale

- 7.3 **We propose to generate exceptions in Model A by comparing the provider and subject-level metrics.** When comparing the metrics, we have chosen to apply a rule based on the initial hypothesis (IH).
- 7.4 **All subjects whose metrics would cause the IH for that subject to be different from the provider-level IH are treated as exceptions.** However, if the change in initial hypothesis is solely as a result of flags changing to neutral at subject-level from positive or negative at provider-level, the subject will not be generated as an exception, as this may be due to a small sample size effect.
- 7.5 Metrics provide an evidence base on which to identify exceptions. However, we do recognise that metrics can only go so far in identifying exception subjects. Given the reliance on metrics and the imperfect nature of the 'hard' rule, we also propose to **allow providers to select a small number of additional subjects to be assessed at subject-level.** These additional subjects would be assessed in the same way as exception subjects. This will allow providers to put forward a subject which they believe deserves greater scrutiny and which, with further evidence, may not match the provider-level rating despite having the same initial hypothesis.

7.6 There are **two options** for identifying the ‘small number’ of additional subjects for assessment:

1. Each provider, regardless of size of provision, is allowed the same number of additional subjects. This could be 1, 2, or 3 additional subjects, to be decided.
2. The number of additional subjects permitted is dependent on the amount of provision (for example, the number of subjects provided, using a formula such as  $n$  subjects divided by 10, rounded up, where  $n$  is the number of subjects at the provider). This would mean that the additional subject allowance was proportionate to the size and range of provision in an institution.

7.7 We are seeking views about which option is preferred. The additional subjects would strictly be in addition to the exceptions generated by the IH rule, not instead of them. All exceptions identified would still require a subject-level assessment.

#### Very high and low absolute values

7.8 **We propose that very high and low absolute values be included in the calculation of the initial hypothesis (IH) when identifying exceptions.**

7.9 As set out in the TEF Specification, where a metric indicator has a very high or low absolute value it will be marked in the metrics workbook with either a star (\*) for very high or an exclamation mark (!) for very low.

7.10 A very high absolute value suggests that a provider’s performance in that metric is so high that, in absolute terms, the experience or outcome for students regarding that metric is outstanding. A very low absolute value suggests that a provider’s performance in that metric is so low that, in absolute terms, the experience or outcome for many students regarding that metric is not good.

7.11 We propose that very high and low absolute values should be considered in a similar way to a positive or negative flag in determining the final position of the IH that is used to identify exceptions. This would follow the existing rules set out in the TEF Specification for applying very high or low absolute values in step 1b of the assessment process. Once very high and low absolute values have been accounted for in both the provider and subject-level metrics, the final position of the resulting IHs will be used to determine whether a subject is an exception.

7.12 If this approach is not taken, this would be likely to create significant inconsistencies between the approach at provider- and subject-level, likely resulting in the identification of ‘exceptions’ which were actually performing in the same way as the provider.

## Alternatives

### Different rules when comparing the metrics

- 7.13 When comparing the provider and subject-level metrics, there are alternative rules that could be considered for determining exception subjects, including:
- **Alternative 1:** Subjects whose metrics differ from the provider-level metrics by at least 1 flag (but not to neutral)
  - **Alternative 2:** Subjects whose metrics differ from the provider-level metrics by at least 2 flags (but not to neutral)
- 7.14 If preferred, the design of these alternatives would need to be worked up in more detail, recognising the complexities, for example, of considering whether marginal cases would arise as a result of these rules, and how they would be treated.
- 7.15 The main disadvantage to Alternative 1 is that our analysis, using TEF data from assessments in academic year 2016/17, suggests that using 1 flag would generate too many subjects to be considered 'exceptions': approximately 54% of subjects across all providers would be exceptions if a difference of 1 flag were used. This would create a huge burden for providers and assessors. It would be impractical and inappropriate for Model A, which aims to reduce the burden on providers and assessors.
- 7.16 Alternative 2 would have a similar amount of exception subjects to the proposed approach, so could be a viable alternative.
- 7.17 The pilots will be testing multiple ways of generating exceptions, including considering these alternatives, but stakeholders are also welcome to provide their view and offer other alternatives when responding to the consultation.

## 8 Model A: Relationship between provider and subject assessment

We are consulting on whether subject ratings should influence provider ratings in Model A. We discuss the option of incorporating a 'feedback loop' into the assessment process.

### Background

- 8.1 We are consulting on whether there should be a stronger relationship between subject and provider assessments and ratings in the design of Model A.
- 8.2 In the current design of Model A, the subject ratings do not influence the provider rating in any way, either for exception or non-exception subjects. The subject assessments for exception subjects occur independently from the provider-level assessments.

### Proposal and rationale

- 8.3 **We are consulting on whether subject assessment should influence provider-level ratings.**
- 8.4 Subject-level TEF is intended to highlight differences across subjects within a provider. For exception subjects, we have identified that they may perform differently to the provider. It is important that subject assessments for exception subjects be focused on the teaching quality and student outcomes for that particular subject, rather than being too heavily influenced by the provider-level rating. For this reason, subject assessments occur independently from the provider-level assessment. Once these assessments have been completed however, there could be a link between subject ratings and the provider rating.
- 8.5 Students will use the provider-level rating to understand the broader context of studying at an institution, so it is important that the rating accurately reflects the teaching quality across all the subjects it provides.
- 8.6 One option we are considering is to incorporate a 'feedback loop' into the process, in which the provider rating would be reconsidered after the subject ratings had been determined. This would mean that if a provider had a larger number of exceptions, and they were consistently moved up or down, the provider rating itself could alter.
- 8.7 One **advantage** of this option is that it would strengthen the holistic character of the assessment process by allowing assessors to take into account all relevant

information when determining the final rating and would allow additional information from the subject ratings to impact the provider-level rating. If the subject ratings made a significant impact on the final provider-level rating, this could be seen as providing students with a more accurate picture of the quality of teaching that an institution provides.

- 8.8 The **disadvantage** of this is that it would add complexity and length to the assessment process and could give undue weight to exception subjects. It might also lead to potential inconsistencies between the provider-level rating and the ratings for non-exception subjects. This could be confusing to students or could be misleading if the final provider-level rating was skewed and therefore did not accurately reflect the broad quality of teaching at that institution.
- 8.9 We welcome views on this, including the option presented and suggestions on how else this could be achieved, if desired.

## 9 Model B: Relationship between provider and subject assessment

In this chapter, we discuss the relationship between the provider and subject assessment in Model B. In this model the subject-level assessment takes place first, and the **provider-level rating is derived from the subject ratings**.

Subject submissions are written for 7 subject groups, but subject ratings are still awarded for the 35 subjects being assessed. The provider-level rating is influenced by the final subject ratings through an element called the 'subject-based initial hypothesis'. This is considered alongside the existing elements of a provider-level assessment. The provider rating is therefore based on a holistic judgement using 3 sources of evidence:

- a) Provider-level metrics – as in provider-level TEF, with an initial hypothesis
- b) Provider-level submission – limited to 10 pages, focusing on 3 criteria
- c) **Subject-based initial hypothesis** –The final subject ratings are weighted by the number of students studying each subject and combined to reach an initial hypothesis for the provider rating.

### Background

- 9.1 In Model B, subject-level assessment is the first stage of assessment and is fully completed before the provider-level assessment takes place. Given this, there is an opportunity for **the provider-level rating to be derived from, or at least influenced by, the subject-level ratings**.
- 9.2 This section is therefore about how the subject-level ratings can feed into the provider-level assessment process.

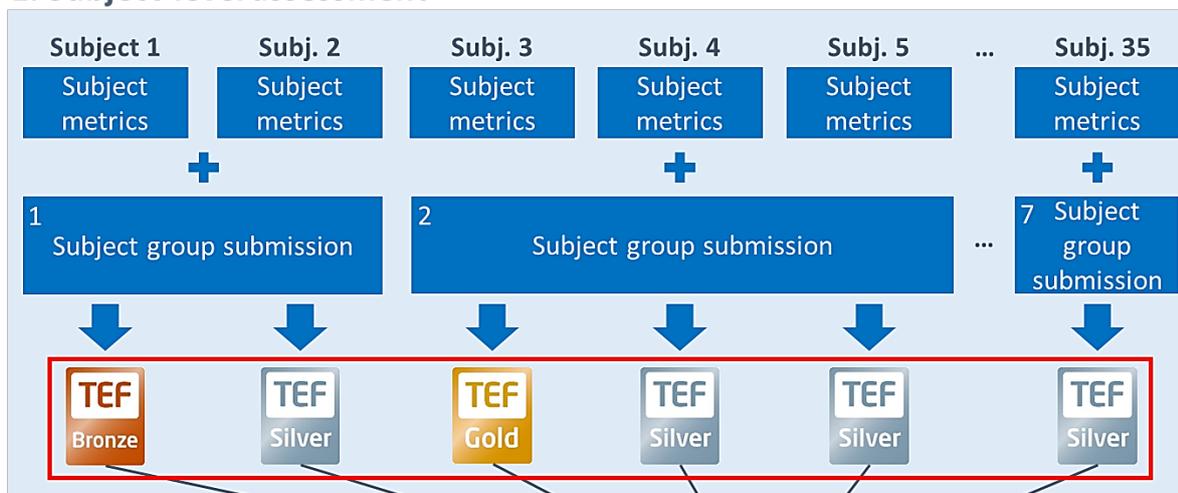
### Proposal and rationale

- 9.3 We propose to assess and rate the provider using 3 sources of evidence:
  - a) **Provider-level metrics, with an initial hypothesis** – generated as in provider-level TEF
  - b) **Provider-level submission** – limited to 10 pages, focusing on 3 criteria
  - c) **Subject-based initial hypothesis** – calculation shown below.
- 9.4 This maintains the existing elements of the provider-level assessment (i.e. the metrics and submission), but adds in a new third element (the subject-based initial hypothesis) that draws on the ratings achieved at subject-level.

- 9.5 The subject-level ratings (generated through full assessment of subject-level metrics and group submissions) are used to form an initial hypothesis of Gold, Silver or Bronze for the provider-level rating. This is called the subject-based initial hypothesis (IH), and will be considered alongside the provider-level metrics and submission in order to reach a holistic judgement about the provider-level rating.
- 9.6 Figure 5 demonstrates how the subject-level ratings feed into the calculation of the subject-based IH.

**Figure 5: Subject-based initial hypothesis in Model B**

**1. Subject-level assessment**



**2. Provider-level assessment**



- 9.7 We held discussions with the sector and the TEF User Group to inform how to weight the influence of each subject rating on the subject-based IH. We propose that each subject rating be weighted by the proportion of the provider’s students that study that subject (calculated by headcount, to remain consistent with the current use of headcount in calculating TEF metrics and benchmarks).
- 9.8 The calculation of the subject-based initial hypothesis will follow the principles used in provider-level TEF, which emphasise a need for consistency in provision in order to do well in TEF. The subject-based initial hypothesis for the provider-level rating in Model B will therefore be calculated by applying the rules in this order:
- If 33% or more of a provider’s students captured in the metrics are in subjects receiving a Bronze rating, the provider will receive a Bronze initial hypothesis.

- If at least 50% of a provider's students captured in the metrics are in subjects receiving a Gold rating, and fewer than 33% are in subjects receiving a Bronze rating, the provider will receive a Gold initial hypothesis.
- All other scenarios generate a Silver initial hypothesis.

9.9 These rules ensure that a provider will receive a Bronze subject-based initial hypothesis if delivering Bronze-rated provision to a significant proportion (a third) of its students – even if most of its provision is rated Silver or Gold. This means that, to receive a Silver or Gold initial hypothesis, a provider must be rated consistently well across its subjects.

9.10 The main panel will consider the subject-based IH alongside the IH based on provider-level core metrics and the provider-level submission in order to reach a holistic judgement and agree a provider-level rating.

## 10 Metrics

This chapter outlines how we propose to apply the existing TEF metrics to subject-level TEF and explores some of the challenges in doing this.

Each participating provider in subject-level TEF will receive both provider-level and subject-level metrics. Subject-level metrics will be reported for each of the 35 subjects in the CAH2 and will include the same metrics and benchmarks as provider-level metrics (except the supplementary grade inflation metric). Providers will have subject-level metrics for each subject they teach.

- **Grade inflation** – the consultation seeks views on whether grade inflation should apply only in the provider-level metrics.
- **Distribution of subject ratings** – We are consulting on whether to allow the distribution of ratings to vary for each subject, or if we should force the same uniform distribution for all subjects. While benchmarking still occurs within subjects, there are certain features of the metrics at subject-level that may affect the distribution profile of subject ratings. These include the marking of very high and low absolute values, clustered metrics, and different levels of external regulation and standardisation between subjects.
- **Non-reportable metrics** – TEF metrics are ‘non-reportable’ if the metric does not meet certain reportability thresholds, such as having at least 10 students contributing to it, or meeting certain response rates. Non-reportability is common for metrics that are based on small sample sizes, so often affects small areas of provision. Non-reportable metrics become more common at subject-level because the metrics are disaggregated and reported separately for each of the 35 subjects. To address this issue we are proposing an alternative approach to treating and assessing subjects with non-reportable core metrics. Under this approach options, some subjects would not be assessed or rated, and where assessment does go ahead, the panel would rely on group- or provider-level metrics.

We note that HEFCE have made institutions’ illustrative subject-level metrics available on the HEFCE extranet. Providers may wish to use this resource to inform their consultation response. Please contact your provider’s TEF contact for details.

## 10.1 Introduction

10.1.1 **The metrics and benchmarks for subject-level TEF will be the same as those used in provider-level TEF.**

10.1.2 Each provider participating in subject-level TEF will receive both:

- provider-level metrics – these will be calculated and reported in the same way as in provider-level TEF
- subject-level metrics – these will be reported for each of the 35 subjects in the CAH2 and will be based on the same suite of metrics as provider-level metrics (excluding grade inflation – see below).

10.1.3 The provider-level metrics are set out in the [TEF Specification](#) published in October 2017. These would remain the same in subject-level TEF. For detailed information about the existing TEF metrics, please refer to the TEF Specification.

10.1.4 Given teaching excellence will be assessed against the same 10 TEF criteria in each subject, the metrics used at provider-level to provide evidence against these criteria remain relevant for all subjects.

10.1.5 In addition to the existing suite of metrics, measures of teaching intensity are also being tested for potential inclusion in subject-level TEF. Teaching intensity is addressed separately in Chapter 13 of the consultation document and this technical document.

10.1.6 The section below sets out how we propose to apply the provider-level metrics at subject level. The grade inflation supplementary metric is addressed separately, as we are proposing to treat this differently at provider and subject-level. The remaining metrics apply in the same way in both provider and subject-level metrics. This follows the overarching principle that the design of subject-level TEF is based on the provider-level framework. This overarching principle is discussed in Chapter 6 of this document.

10.1.7 In the main, the analysis, which has informed the consultation, uses TEF data from assessments undertaken in academic year 2016/17 ('TEF Year Two'). HEFCE have provided some high level information based on TEF data prepared for assessments in academic year 2017/18 ('TEF Year Three'), but we did not have access to the full anonymised data set for this year when this consultation was written. DfE does not obtain TEF data until the application window closes in order to maintain independence from the assessment process.

## Subject-level metrics

### Core and split metrics

10.1.8 The criteria for assessing providers is also relevant for assessing subjects.

Therefore, the core metrics and associated splits also apply at subject-level, and subject-level metrics will be calculated in the same way as provider-level metrics. All core metrics and splits will be reported separately for full-time and part-time students.

10.1.9 The initial hypothesis in subject-level metrics will also be calculated in the same way as in provider-level TEF.

10.1.10 We are not proposing to make any changes to the core metrics or splits when applying these to each subject. If there are any issues about how the core metrics or splits apply for a particular subject area, we welcome feedback on these and suggestions for how such issues could be resolved (e.g. through adjustments to the metrics or additional metrics that may help to resolve the issue for a specific subject).

### Benchmarking

10.1.11 We are proposing that subject-level metrics be benchmarked in the same way as provider-level metrics. We are proposing to use the same benchmarking factors and groupings as at provider-level, and to apply these consistently across all 35 subjects. An alternative would be to consider having different benchmarking factors for each subject. This will be explored through the evaluation of the pilots.

10.1.12 Benchmarking will be undertaken within each subject, using data from across the higher education sector. 'The sector' will be made up of all providers who report students in the given subject area in a HESA data return, regardless of their participation/eligibility in provider or subject-level TEF. Benchmarking within subjects ensures that benchmarks account for any differences in outcomes that may be driven by choice of subject, rather than by institutional performance, meaning that the assessment focuses on teaching excellence within each particular subject. From a student perspective, this means that subject-level TEF will allow students to compare different providers offering the subject they wish to study, but will not attempt to make value judgements as to the relative worth of different subjects.

10.1.13 The approach to flagging based on benchmarks in the subject-level metrics will be the same as for provider-level TEF.

## Supplementary metrics based on Longitudinal Education Outcomes data

- 10.1.14 Subject-level metrics will include the Longitudinal Education Outcomes<sup>2</sup> (LEO) supplementary metrics and these will be used in the same way as for provider-level TEF (i.e. alongside submissions).
- 10.1.15 Both supplementary LEO metrics will be defined in the same way as for provider-level TEF. However, to ensure that no provider is able to identify any individual student's contribution (or lack thereof) to the supplementary LEO-based metrics, an additional suppression has been applied to these metrics at subject level. See HEFCE's [TEF subject pilot metrics: specification and rebuild document](#) for further information.
- 10.1.16 For the salary LEO metric, this means that the median salary threshold will remain at £21,000, as per provider-level metrics. The current threshold of £21,000<sup>3</sup> is based on the median salary for taxpayers aged 25-29 and so provides an indication of how likely it is that a student's investment in obtaining a degree in that subject will lead to an above average level of earnings.
- 10.1.17 All metrics, including those based on LEO, are benchmarked by subject so as to assess performance within each subject. This means that a provider's subject-level LEO metrics will be compared to a benchmark that is specific to each subject.
- 10.1.18 An alternative approach could be to apply subject-specific median salary thresholds when calculating this LEO metric for each subject. This would explicitly recognise that graduates' earning outcomes are likely to systematically differ by subject in a way that is unrelated to teaching quality. However, this approach would add significant complexity to the assessment process and potentially prove confusing for students, with employment performance no longer measured on the same absolute basis and without reference to general earnings levels in the economy (there is no equivalent measure on a subject basis). Furthermore, it is unnecessary, as benchmarking already takes place within each subject, meaning that subjects are not being compared against each other, but only to the same subject at other providers. For these reasons, we are not proposing to use subject specific median earnings thresholds.

---

<sup>2</sup> <https://www.gov.uk/government/statistics/graduate-outcomes-leo-including-self-employment-earnings-data>

<sup>3</sup> Rounded to the nearest £500. This value will be updated annually in accordance with the relevant dataset.

## 10.2 Grade inflation

10.2.1 We are proposing that the grade inflation metric would only be included in the provider-level metrics and not in the subject-level metrics. This is because we understand that decisions around grade boundaries are often set at an institutional level covering the whole provider. Disaggregating grade inflation data to subject level would also see an increase in the small sample size effects and natural variation from year-to-year, which will be stronger at subject-level, making the metric less robust.

10.2.2 We invite feedback from providers about whether they agree with this approach and whether the rationale applies consistently across the sector.

## 10.3 Distribution of subject ratings

### Background

10.3.1 The **distribution of final ratings is unlikely to be the same in each subject**. For some subjects, the subject-level metrics may be quite similar across providers and we may see a high proportion of providers achieving, for example, higher TEF ratings of Silver and Gold. Conversely, we could see some subjects with a higher proportion of Bronze ratings. For some other subjects, there may be more variation in the metrics and we may see providers more widely spread across Bronze, Silver and Gold ratings.

10.3.2 Benchmarking occurs at a subject level, ensuring fairer comparisons of provider performance by avoiding biases that could arise if subject-specific expectations were not factored in. However, this does not mean that we would expect each subject to produce the same distribution of ratings. Reasons for this include different patterns of very high and low absolute values; the extent to which metrics are clustered in each subject; and different levels of external regulation and standardisation between subjects, which could impact both the metrics and the degree of variation between provider submissions.

### Proposal and rationale

10.3.3 Under the proposed design of subject-level TEF and the subject-level metrics, we expect that different subjects may naturally have different rating distribution profiles. We are proposing to allow this effect to occur, meaning we will **allow the distribution of subject ratings to vary naturally for each subject, rather than forcing a uniform distribution**.

10.3.4 We believe allowing rating profiles to vary naturally by subject is an appropriate outcome because it rewards excellence where it is found and does not prejudge

the 'value' of one subject compared to another. This will give students clear information about where the best teaching and outcomes can be found for each subject:

- Individually, if a provider performs well against the metrics and the evidence suggests they are offering that subject to a Gold or Silver standard, then it is appropriate to signal this to students by awarding that provider a Gold or Silver rating respectively. Similarly, if a provider is offering a subject at a Bronze standard, this should be signalled to students through a Bronze rating.
- Across all providers offering a subject, if a high proportion of them offer that subject at a Silver or Gold standard, then it is similarly appropriate to signal this to students by awarding a high proportion of Silver or Gold ratings for that subject. Similarly, if a large proportion offer Bronze standard, this should be signalled to students.

10.3.5 From a student information perspective, if a large number of providers get the same rating for a particular subject, this would provide less differentiation between providers for that subject. Differentiation will not be removed entirely, however, as we expect it would be unlikely for 100% of providers in a particular subject to achieve the same rating. Therefore, the ratings should still tell a student which providers are offering the best teaching quality and where the best outcomes are on offer. If a large proportion of providers are rated Gold for example, then this informs students that if they want to study that subject, they can access a similarly high level of quality and student outcomes at a large number of institutions across the sector. Equally, by allowing natural variation in distribution ratings the 'value' of one subject compared to others is not prejudged. This will still give students clear information about where the best teaching and outcomes can be found for each subject.

10.3.6 Through the pilot, we will test how the distribution of final subject ratings differs across subjects.

#### **Alternative: Forcing a uniform distribution with quotas**

10.3.7 The alternative to our proposal would be to set quotas that would create a uniform distribution for subject ratings. This would mean that each subject would have the same percentage of Gold, Silver and Bronze awards. When assessors are moderating for consistency of how ratings are awarded, this would mean that ratings for each subject were moderated within the subject, rather than performing moderation across all subjects.

10.3.8 This alternative would depart from the robust evidence-based assessment process that TEF is currently built on. If this approach was taken, panels would

be forced to move ratings up and down to fit the uniform distribution. This could result in a provider being awarded a rating where their performance is not a best fit or match to the descriptor for their award.

10.3.9 We do not believe this is an appropriate approach, as it would not provide a fair and robust assessment of subject provision.

### **Impact of very high and low absolute values**

10.3.10 The approach to highlighting very high and low absolute values in the subject-level metrics may affect the extent to which there are differences in the distribution of subject ratings.

10.3.11 We have considered two options for identifying very high and low absolute values in the subject-level metrics. Our proposed option would potentially have a larger impact on the distribution of ratings across subjects than the alternative option presented. However, final ratings are made based on a holistic assessment process, and so the actual impact on subject ratings is still uncertain at this stage and will be tested in the pilots.

### **Calculation in provider-level metrics**

10.3.12 As set out in the TEF Specification, in provider-level TEF metrics, where a core TEF metric or a supplementary LEO metric has an indicator with a very high or low absolute value it will be marked in the metrics workbook. To identify values that are defined as very high or very low, a threshold is calculated.<sup>4</sup> This threshold then applies to all providers, meaning all providers with an absolute value above the identified very high threshold receive a star (\*), and any with an absolute value below the identified very low threshold receive an exclamation mark (!).

10.3.13 Whilst benchmarking still remains at the heart of TEF assessments, the flagging and benchmarking system does have limitations at the extreme ends of the metrics, where it is harder for a metric to be flagged. Marking very high and low absolute values addresses this limitation and ensures that these absolute measures are sufficiently accounted for in the assessment process.

10.3.14 From a student perspective, accounting for absolute values is important because we want TEF to recognise good student outcomes where they occur. For example, if satisfaction rates on Assessment and Feedback, or Progression to Highly Skilled Employment and Further Study are very high, this is what will matter most to an individual student. Students may also be concerned with very

---

<sup>4</sup> The thresholds at provider-level are based on the top and bottom 10 per cent of providers (for the metrics and mode in question) that have a reportable metric that refers to more than 100 students.

low absolute results where, even if the provider is on benchmark, in practice a large number of students will not be receiving positive outcomes.

#### Application to subject-level metrics

10.3.15 Very high and low absolute values are used in TEF in recognition of the fact that (a) due to the law of diminishing returns, it may be harder to achieve material improvements when performance is already very high and (b) from a student perspective, a very high (or low) absolute performance represents a positive (or negative) outcome, and should be recognised as such. These principles apply equally at subject-level. The flagging and benchmarking approach in the subject-level metrics has the same limitations at the extreme ends of the metrics. The importance of recognising very high and low outcomes for students is also equally important for subject assessment as it is to provider-level assessment.

10.3.16 Therefore, we propose to mark very high and low absolute values in the subject-level metrics in same way as they are marked in provider level metrics. They will also be used in the subject assessment process in the same way as provider-level TEF (at Step 1b – for more information on this, please see the TEF Specification).

10.3.17 To identify very high and low absolute values in the subject-level metrics, we have considered two alternative options:

1. Apply the **same thresholds** in subject-level metrics that are already defined for provider-level metrics.
2. Apply **different thresholds** for each subject in subject-level metrics, calculating these using the same approach as provider-level thresholds, which would focus on the top and bottom 10 percent within a subject. For each subject, the threshold would be defined separately for each metric and mode of study.

10.3.18 The threshold values under these two options are shown in ‘Annex B: Absolute value thresholds’.

10.3.19 The key difference between these approaches is whether the very highest or lowest performance should be assessed with respect to sector-wide outcomes or with outcomes specific to its subject. On the one hand, comparing against sector-wide outcomes provides a clear signal to students about what should be considered normal and exceptional in their degree experiences and outcomes. On the other hand, there may be reasons unconnected to teaching quality that are linked to particularly strong or weak outcomes in a subject and so, in an extreme circumstance, may potentially mask the teaching quality being provided.

10.3.20 **We are proposing to use Option 1 (same thresholds)** in the subject-level metrics. We are testing this option in the first year of pilots and this is reflected in the subject-level metrics that HEFCE have made available to providers.

#### Rationale for proposing Option 1: Same thresholds

10.3.21 We have given a preference to option 1 because the provider-level thresholds represent how very high and low absolute values have already been defined and because, by definition, absolute values should be about absolute, not relative, performance (the latter being addressed by benchmarks). Under this approach, there would be consistency in the definition of very high and low absolute values. If a value were defined as very high or very low at provider-level, it would also be defined as very high or very low in the subject-level metrics. This should aid both student and provider understanding of the process, and make subject-level TEF more accessible for both students and providers. The evaluation of the pilot however, will look to explore the sensitivity of results to this aspect of the design.

10.3.22 One **implication of option 1** is that it may have a stronger impact on the distribution of ratings across subjects such that different subjects have different rating distribution profiles.

10.3.23 Under option 1, the subject-level metrics will still be benchmarked within each subject, so consideration of relative performance within a subject would still be a key component of the assessment process.

#### Alternative

10.3.24 The **alternative approach** (option 2 – different thresholds for each subject) would incorporate a further element of relative performance within subjects into the assessment process. This would potentially complement the existing benchmarking approach by subject, meaning that subject-level assessments would be based purely on performance within each subject.

10.3.25 The **key disadvantage of option 2** is that it offers students perverse information about where they might expect to see very high or very low absolute values. Given the variability in metric performance across subjects, using subject specific thresholds could result in, for some subjects:

- Very high thresholds that are actually not that high in absolute terms;
- Very low thresholds that are actually quite high in absolute terms.

10.3.26 For students, option 2 means that a TEF rating would not be informed by a consistent approach to measuring absolute performance. For some subjects, a provider's performance may not receive a star (\*), or could actually receive an

exclamation mark (!) even though they are achieving good outcomes for almost all of their students.

10.3.27 For example, at the top end of the metrics, under option 2, the very low absolute value threshold for the 'employment or further study' metric for full-time students for Medicine & dentistry would be 99.41%<sup>5</sup>. This means providers offering this subject could be marked down as very low performance even though 99% of their students achieved a positive employment outcome. From a student perspective, it seems sensible that this high level of outcomes should be seen as a positive outcome rather than a neutral or potentially even a negative one.

10.3.28 The effect is particularly prevalent for subjects such as Medicine & dentistry and Nursing, where most or all providers have very high absolute values for certain metrics for full-time study (see clustered metrics section below). In this case, applying subject thresholds would disadvantage providers who are not in the top 10% of that subject, even though they have performed well in absolute terms when defined across all subjects (i.e. at provider-level). Given this effect, we do not believe this would fairly assess providers in terms of absolute performance.

10.3.29 Therefore, under option 2 (subject-specific thresholds), absolute performance would not be treated consistently across subjects.

10.3.30 It is also worth noting that sample sizes become too small to calculate subject-specific thresholds robustly for part-time provision for each subject. The problem of smaller sample sizes is a challenge of option 2 that would need to be addressed if this option was taken forward.

#### Impact on the distribution of subject ratings

10.3.31 The impact of both options on the final subject ratings achieved by providers is still unknown at this stage. This is because the final rating is based on a holistic assessment, which includes not only metrics, but also consideration of subject submissions. The impact on final ratings of option 1 will be tested through the first year of the subject-level pilots.

10.3.32 We welcome views from stakeholders about whether option 1 or 2 is preferred and the implications of each approach.

#### Clustered metrics

10.3.33 For some subjects, the metrics of providers offering that subject are clustered together within a very small range, often but not always at the top end

---

<sup>5</sup> This is based on subject specific thresholds produced by HEFCE based on TEF data from assessment undertaken in academic year 2017/18. Please see Annex B for a table of the highest and lowest subject specific thresholds for each metric under Option 2.

of the spectrum. Subject-level TEF predominantly assesses performance within each subject by benchmarking the metrics by subject. However, for some subjects, the clustering effect impacts on how metrics are flagged. At the top end of the spectrum, it also impacts on how they are marked as very high absolute values. These factors may result in differences in performance across subjects. Therefore, for subjects with clustered metrics, we may see many providers achieving higher TEF ratings of Silver, and at the top of end of the spectrum, Gold.

10.3.34 We can see an example of clustered metrics when looking at the TEF data for assessments in academic year 2016/17 for the ‘employment and further study’ and ‘highly skilled employment or further study’ metrics. For these metrics, the clustering effect can be seen consistently across both full-time and part-time provision for two subjects:

- Medicine & dentistry
- Nursing.

10.3.35 The strongest effect is seen in Medicine and dentistry, where the range of providers’ indicators is significantly smaller than in every other subject, and the median of the indicators is consistently higher.

10.3.36 For these subjects, the clustering effect occurs at the top end of the spectrum. It is likely that clustered metrics at high values occur because these subjects are highly regulated and/or have widely adopted accreditation schemes with high quality standards. They also have more clearly defined employment paths, some of which may be subject to employment market shortages (e.g. healthcare professionals) and higher education may be the only route into this employment. This means students often go straight into highly skilled employment after graduating.

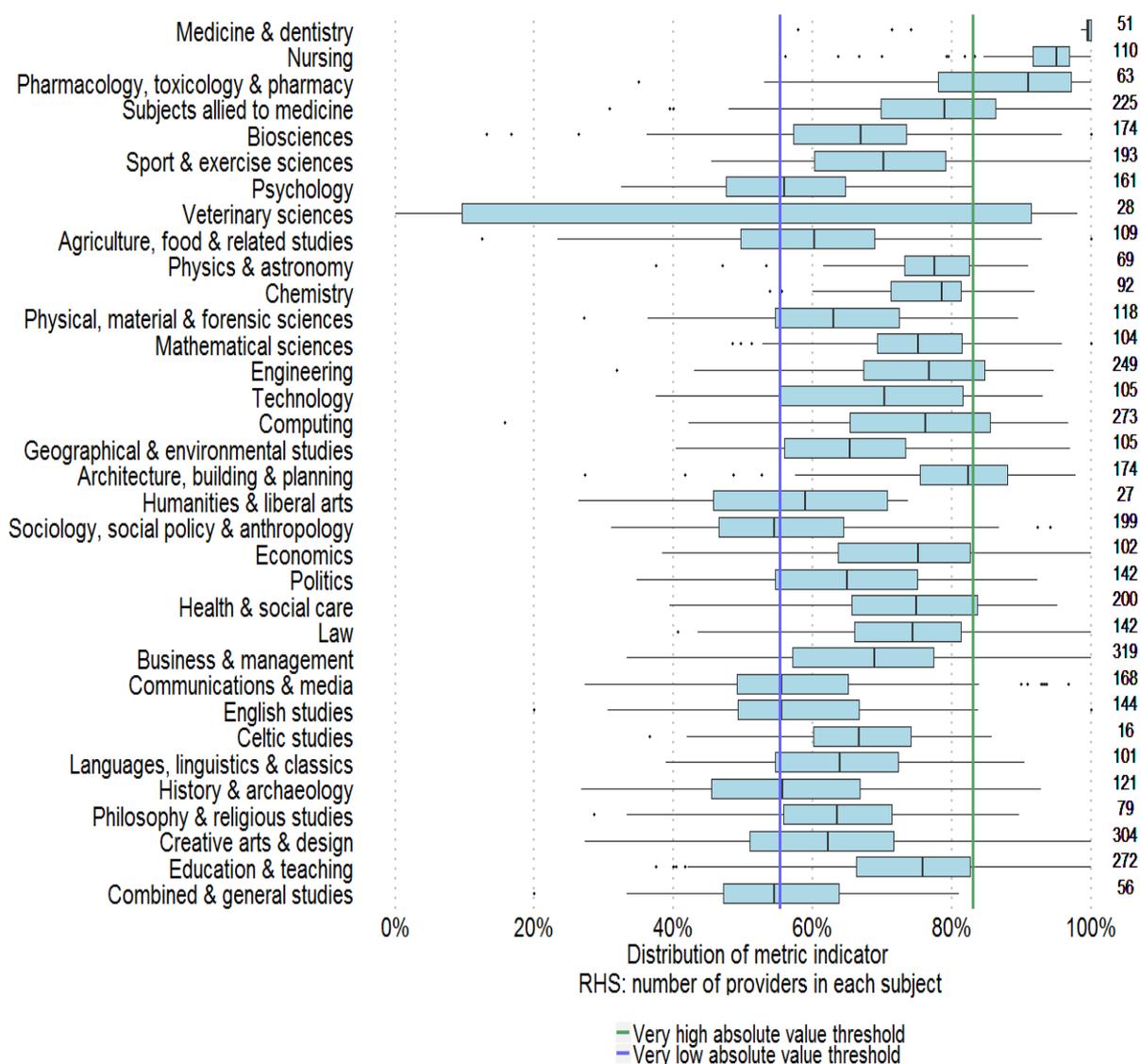
10.3.37 Figure 6 shows a box plot of the highly skilled employment or further study metric for full-time students across all providers, showing how the range of indicators compare for each of the 35 CAH2 subjects. This is based on TEF data for assessments in academic year 2016/17<sup>6</sup>. This graph shows the clustering effect for Medicine & dentistry and Nursing. The right hand side shows the number of providers that offer that subject which range from 16 providers for Celtic studies to 319 providers for Business and management. The majority of

---

<sup>6</sup> This includes all providers that were in scope for TEF during the TEF assessments that took place in academic year 2016/17. It should be noted that this includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year.

subjects had at least 100 providers. Similar box plots for all six metrics are provided in 'Annex C: Metrics charts', for full time students.

**Figure 6: Distribution of highly skilled employment or further studies indicators (full-time)**



**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low thresholds shown on the graph have been provided by HEFCE, based on TEF data prepared for assessments in academic year 2017/18. The chart shows the very high absolute value threshold (83.01%) and the very low absolute value threshold (55.30%). Each boxplot shows the overall pattern of a metric indicator across providers for a subject. The mid-point of the data is shown by the line that divides the box into two parts. Half the indicator values are greater than or equal to this value and half are less. The length of the box or the whiskers indicates how varied the indicator is across providers for a subject, and the black dots represent outlying values. For example, for Veterinary sciences, there is little variation across providers at the top and bottom (shown by the short whiskers) but there is a lot of variation in the middle (shown by the long box). In contrast, for nursing, there is little variation across providers.

10.3.38 The TEF data for assessments in academic year 2016/17 also shows clustered metrics for a small number of other subjects, but the effect is less consistent across the metrics and varies across full-time and part-time provision. The clustering effect is seen at the top and middle ranges of the metrics, but is not seen at the lower end of the spectrum for these data. Please see Annex C: Metrics charts for further charts.

#### Impact on the distribution of subject ratings

10.3.39 Where metrics are clustered together, this affects the performance of that subject in the assessment process in two ways:

- **Flags** – It becomes less likely that these metrics will be flagged because the clustering means it is less likely that a provider will differ from the benchmark by more than the materiality test (at least 2 percentage points away from benchmark). This does not apply at the very top end of the spectrum however, as the materiality test does not apply if the metric indicator is above 97% and this is above benchmark.
- **Very high absolute values** – If the clustering is very high or very low, it may also mean that more providers in that subject, for that metric, will receive a star (\*) or exclamation mark (!) for very high and low absolute values.

10.3.40 These effects may impact on the initial hypothesis and subsequent rating. Due to the effect on flags, it is possible that subjects with clustered metrics may have a higher proportion of providers with a Silver initial hypothesis. Where the metrics cluster at the top end of the spectrum, the effect on very high absolute values may offset this and move these initial hypotheses to Gold. Further analysis will be undertaken in the pilot to test this in the subject-level metrics.

10.3.41 On balance, we expect that when the impact of clustering on flags and very high absolute values is accounted for (using provider-level thresholds), subjects with clustered high values will tend to perform particularly well in the metrics part of the assessment process. For these subjects therefore, **we may see a higher proportion of providers achieving Gold and Silver awards compared to other subjects**. This will be tested in the pilots.

#### Regulation and standardisation

10.3.42 Different subjects are placed under different requirements regarding regulation and standardisation by Professional, Statutory and Regulatory bodies. This may indirectly impact on the metrics (if similar styles of teaching are adopted, the outcomes may be more similar) and on the provider submissions, as providers may be adopting similar practices. This in turn may result in more providers getting the same rating, whether that is Bronze, Silver or Gold.

## 10.4 Non-reportable metrics

### Background

10.4.1 For TEF metrics to be reported and used in the assessment process, they must meet certain reportability thresholds. The current thresholds are:

- 10 or more students are referred to by the metric<sup>7</sup>
- Response rate of 50% for the National Student Survey (NSS)
- Response rate of 85% of the target response rate for the Destination of Leavers from Higher Education survey (DLHE) (85% of the target population, which is equivalent to 68% for full-time students and 59.5% for part-time students)
- There is sufficient data to form a benchmark for that metric<sup>8</sup>
- For the LEO supplementary metrics:
  - a numerator that differs from the denominator by at least three students
  - in calculating the earnings metric, at least 50% of those in the denominator have earnings data or further study information.

10.4.2 Eligibility to take part in TEF will continue to be based on a provider having a suitable set of provider-level metrics. Non-reportable metrics at subject-level will not affect a provider's eligibility for TEF, but will affect how those subjects are assessed.

10.4.3 At subject-level, metrics will be broken down and reported separately for each of the 35 CAH2 subjects. Once the metrics are disaggregated at this level, the sample size of students used to calculate each metric is smaller and it becomes harder to meet each of the reportability thresholds. This means most providers will find that some of their metrics for some subjects become non-reportable.

10.4.4 Non-reportable metrics may result in two issues for the assessment process:

- **For core metrics**, some providers would not meet the 'suitable metrics' thresholds for some subjects. This means there would be insufficient data to calculate an initial hypothesis (IH) for that subject, which should be calculated using all 6 core metrics.

---

<sup>7</sup> There has to also be at least 10 students in the numerator, otherwise the indicator is suppressed for data protection reasons.

<sup>8</sup> Sufficient benchmarking data would be at least 50% coverage for each factor (for example where entry qualifications are used as a benchmarking factor, at least 50% of the provider's students included in the core or split metric must have appropriately recorded entry qualifications).

- **For split and supplementary metrics**, this means there may be no data or very limited data available to inform refinements to the IH in step 1b of the assessment process.

10.4.5 While metrics are not the only source of information in TEF assessments, they do form an important part of giving assessors sufficient information to make a fully informed and robust judgement. Specifically, core and split metrics are used explicitly in assessment parts 1a and 1b to calculate the IH.

10.4.6 Given the metrics used in assessment are the same in both Model A and B, non-reportable metrics are an issue under both models. However, the way in which they affect assessment varies. This is outlined below in discussing how the options would apply to each model.

### Eligibility under subject-level TEF

10.4.7 **Eligibility requirements in relation to suitable metrics for participating in subject-level TEF will be based on the current provider-level requirements.**

10.4.8 To ensure that, as it moves to subject-level, the TEF continues to recognise the diversity of the sector, we want to ensure that providers who meet the eligibility requirements for provider-level TEF, and have suitable metrics at provider-level, continue to be eligible to participate in the TEF once it moves to subject-level, even if not all of the subjects it has on offer have a full set of reportable core metrics at subject-level.

10.4.9 As a result, eligibility for full assessment under subject-level TEF will continue to be based on provider-level metrics. That is, providers must have a minimum set of reportable core metrics at provider-level in order to apply for assessment in subject-level TEF.

10.4.10 Under subject-level TEF, if a provider does not meet the suitable metrics requirement in their provider-level metrics, they will still be able to opt in for a provisional provider-level award.

### Core metrics

10.4.11 In provider-level TEF, a provider is required to have 'suitable metrics' at provider-level in order to be eligible for a full TEF assessment. This means having one year of reportable, benchmarked data for **each of the six core metrics**, for either full- or part-time students, whichever forms the majority taught at the provider. At provider-level, providers that do not have suitable metrics may opt-in for a provisional TEF award. This would continue to apply under subject-level TEF.

10.4.12 At subject-level, given the increase in non-reportable metrics, this 'suitable metrics' threshold may not be met for all subjects, even if it is met at provider-

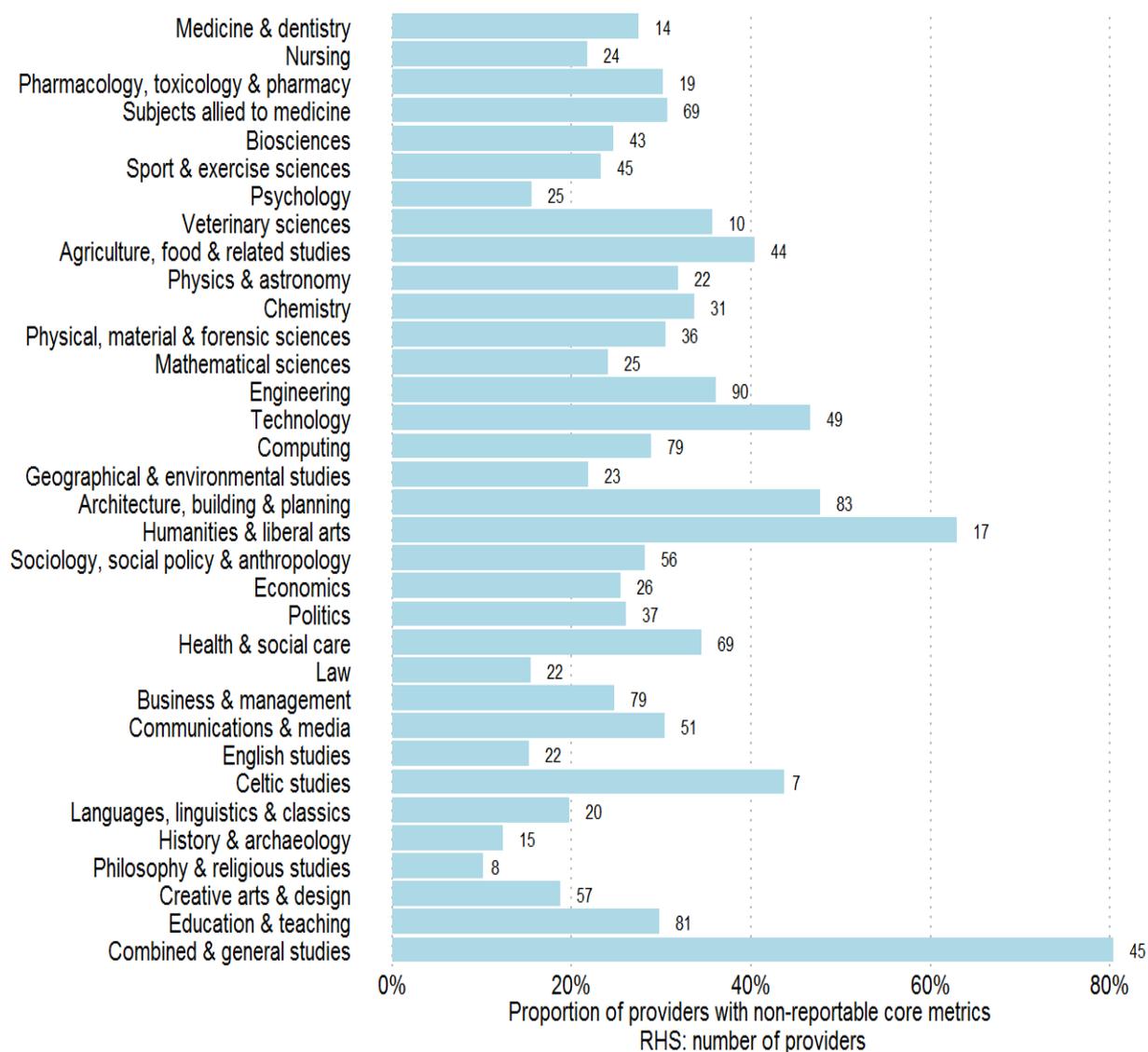
level. Analysis based on TEF data from assessments in academic year 2016/17 suggests that although 98% of students will still be in subjects with suitable metrics, 87% of providers would have non-reportable core metrics in at least one subject (because most providers have at least one small subject). This is particularly the case for alternative providers and colleges, which often have smaller cohorts.

10.4.13 The graph shown in Figure 7 shows the number and proportion of providers that had non-reportable core metrics (do not have 'suitable metrics') for each of the 35 subjects using TEF data for assessments undertaken in academic year 2016/17<sup>9</sup>. The right hand side shows how many providers had non-reportable core metrics for that subject. The graph shows that in all subjects some providers had non-reportable core metrics, but that the extent of the problem varies across subjects. It was most prevalent in two of the 'general' subjects – 'Combined and general studies' and 'Humanities and liberal arts' with over 80% (45 providers) and 60% (17 providers) of providers, respectively, who had non-reportable core metrics in these subject areas. Subjects that had a low percentage of providers with non-reportable core metrics include Philosophy and religious studies (8 providers), History and archaeology (15 providers), English Studies (22 providers), Law (22 providers) and Psychology 25 providers).

---

<sup>9</sup> See notes to Figure 7 for further information.

**Figure 7: Number and percentage of providers that have non-reportable core metrics**



**Note:** Analysis is based on TEF data for assessments undertaken in academic year 2016/17 which includes all providers that were in scope for TEF during the TEF assessments that took place in academic year 2016/17. This includes data from providers who were eligible for full TEF awards, and differs from the number of providers who received a TEF award in that assessment year. A provider is counted as having non-reportable metrics if they do not have all six core metrics by majority mode.

10.4.14 Table 3 below shows how the extent of non-reportable metrics differs by provider type. The table shows the proportion of subject instances (the number of individual subjects being assessed across all providers) that had non-reportable core metrics (by majority mode for each subject) and the proportion of students within these subject instances, by provider type. Subject instances are identified by looking at the number of subjects offered by each provider and adding this up across the sector. For example, if provider 1 offers 3 subjects and provider 2 offers 5 subjects, then there are 8 subject instances in total. It is useful to consider subject instances, as it allows us to see these in terms of how many subjects would be assessed.

10.4.15 Non-reportable metrics were significantly more common for Further Education Colleges and Alternative Providers compared with Higher Education Institutes (HEIs). HEIs had the least number of subject instances with non-reportable core metrics (13%), however this represented 126 providers. Further Education Colleges had a large proportion of subject instances with non-reportable core metrics (85%), which represented 177 providers, and alternative providers had about half of their subject instances with non-reportable metrics, but this represented only 5 providers. Although the proportion of subject instances with non-reportable core metrics varies considerably by provider type, only a very small proportion of all students were impacted in each case.

**Table 3: Proportion of non-reportable core metrics (by majority mode) by type of provider**

Provider type	Proportion of subject instances with non-reportable core metrics	Proportion of all students in these subjects
Higher Education Institutes (HEI)	13% (126 providers)	1.2%
Further Education Colleges (FEC)	85% (177 providers)	0.8%
Alternative providers (AP)	50% (5 providers)	0.06%

**Notes:** Analysis is based on TEF data from assessments undertaken in academic year 2016/17 which includes all providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full TEF awards, and differs from the number of providers who received a TEF award in that assessment year. A subject instance is counted as having non-reportable metrics if it does not have all six core metrics by majority mode. For students whose course subjects span more than one CAH level 2 subject, their contribution is split across the two according to the broad balance of subjects i.e. a Maths and Modern Dance student would typically count as a half in each of mathematics and creative arts and design.

## Proposal and rationale

### Split and supplementary metrics

10.4.16 **We are not proposing any restriction or change to the assessment process as a result of split and/or supplementary metrics being non-reportable.**

10.4.17 At provider-level, split and supplementary metrics do not form part of the eligibility requirements for a TEF assessment and we propose to carry forward this principle for subject-level assessments.

10.4.18 Split and supplementary metrics will always be displayed if the provider has them. If they are not reportable, the assessment will continue without them. **Panellists and assessors should not allow splits that do not display flags to affect their hypothesis.**

## Core metrics

10.4.19 If a subject has non-reportable core metrics (i.e. it does not have 'suitable metrics'), the assessment process would be less robust because assessors would have insufficient information available to make a fully informed judgement about that subject. Specifically, it would not be possible to calculate an IH in the same way as worked out at provider-level. The judgement would also be less comparable to other TEF assessments, as the evidence drawn on by assessors would not be the same (i.e. the core metrics component would be missing).

10.4.20 **Given this, we believe the existing assessment process should be adjusted from its current form when a subject has non-reportable core metrics.**

10.4.21 There are several approaches that could be taken to adjust the assessment process where non-reportable core metrics exist. At this stage, we are proposing a particular approach and are seeking views from stakeholders about whether this is most preferable. We also welcome new ideas and other options that stakeholders would like us to consider.

10.4.22 The proposed approach for treating subjects with non-reportable core metrics combines three options, which are summarised below. Each option is explained in further detail including an assessment of the advantages and disadvantages of the option later in this chapter.

### Changes to eligibility for assessment of a subject

1. **Lower suitable metrics threshold** – Apply a more lenient definition of 'suitable metrics' that requires less than all 6 core metrics to be reportable. A different assessment process could then be applied for different levels of non-reportability.

### Changes to the assessment process for a subject

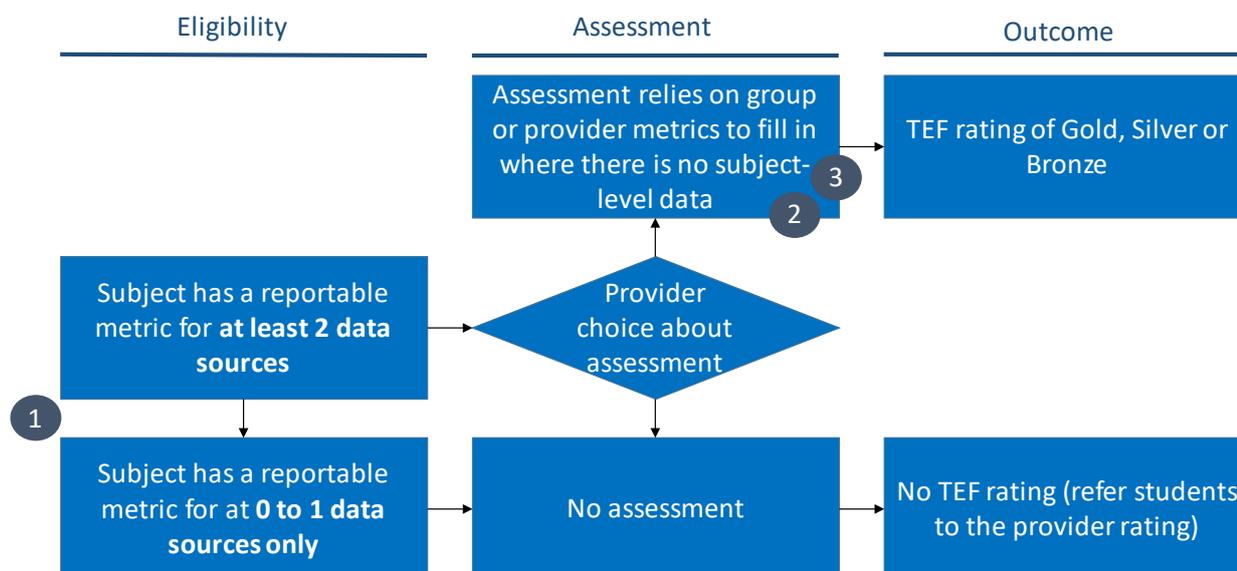
2. **Group metrics** – Rely on metrics reported for the 7 subject groups (the groups used for submissions in Model B) to inform the assessment of that subject. These would be used alongside any reportable subject-level metrics.
3. **Provider metrics** – Rely on the provider-level metrics to inform the assessment of that subject. These would be used alongside any reportable subject-level metrics.

10.4.23 **These options are not mutually exclusive.** Our proposed approach combines these options together.

10.4.24 The diagram in Figure 8 demonstrates our proposed approach to bringing together the options described above. The numbers in the diagram correspond to the list of options outlined above. This proposed approach is based on the

principle of maintaining a sufficient level of metrics within the assessment process. For an assessment to go ahead and a rating to be awarded for a subject, we believe that the lower suitable metrics threshold set in option 1 must be met. That is, the subject must have a reportable metric for at least two data sources – see section below on option 1. We currently have no proposed preference between options 2 and 3 and are seeking views on this.

**Figure 8: Proposed approach to address non-reportable core metrics**



## Alternatives

10.4.25 There are a series of other alternatives that we have also considered to address this issue. We do not think these have sufficient merit because their disadvantages are too great. These are:

- a) Increase weight on the submission / reverse the process – The assessment process would not be sufficiently robust to make a well-informed judgement and the assessment and ratings would not be comparable to other TEF ratings.
- b) Alternative a, but put a marker on the ratings to communicate that they are less robust and indicative only – This would not provide clear information to students and may be confusing or misleading.
- c) Rely on other data submitted by the institution – These data would not be consistent across providers and it may not be subject to the same quality assurance as TEF metrics. Placing heavy reliance on this data may therefore make the assessment process less robust.
- d) Lower reportability thresholds – These thresholds are included in the TEF to ensure the assessment process is based on statistically robust data.

We want to ensure that we maintain the quality of the data being used in TEF metrics.

- e) Aggregate data over more than 3 years – This would lower the responsiveness of TEF assessments and ratings to improvements in teaching excellence and student outcomes.
- f) Produce only a statement of findings – Even if a rating is not awarded, the assessment process would still not be sufficiently robust to make a well-informed judgement about what to write in the statement of findings.
- g) Award a TEF ‘accredited’ or provisional award – Early feedback suggests this approach is unpopular with the sector as it might be misleading or confusing for applicants and students.

10.4.26 Another alternative we considered was to set a **student cohort threshold** (for example, 30 students captured by the metrics for that subject). If a subject only has a small number of students (i.e. below a certain student cohort threshold), assessment could become optional and providers can choose not to be assessed for that subject.

10.4.27 While this option could be viable, it is not included in our proposed approach, as it would add an additional layer of complexity that may not be necessary. However, as we deem this a viable option, we have outlined how this could work and its key advantages and disadvantages below.

10.4.28 Under this alternative option, if a subject has less students in its cohort than the threshold, the provider could choose not to have that subject assessed. That subject would not receive a TEF rating. For all other subjects above the threshold, this choice would not be available and they would be assessed where required.

10.4.29 Where no rating was given under this alternative, there are two proposed approaches for how this could be presented to students:

- a) No rating or reference to TEF at all for that subject (in the same way that a subject not taught by the provider would have no TEF rating)
- b) A TEF statement listed next to that subject that informs students that the subject was too small to be assessed under TEF, and that students may wish to refer to the provider-level rating instead.

10.4.30 This would not be a standalone option. The assessment of a subject which falls below the threshold would need to be informed by one of the options for changing the assessment process (i.e. option 2 or 3 above).

- 10.4.31 This option may be **sensible** to account for the following scenarios:
- The provider does not perceive the subject in question to be one of the major subjects taught by the institution, but they still have a small number of students mapped to that subject in the metrics.
  - The subject is no longer taught at the institution or the subject is being phased out and new students are no longer being recruited for the subject, but historical data is still captured in the metrics.
- 10.4.32 There are **two key disadvantages** of this alternative option. Firstly, there could be a perverse incentive for some providers to avoid assessment of their small subjects if they think these subjects will not perform well in the TEF. Secondly, this option would provide less information to students, with some subjects having no TEF information even though there is a clear (albeit small) cohort of students seeking to study that subject at that institution. The absence of a rating for these subjects could also be perceived negatively by these students.

## The proposed approach

### Option 1: Lower suitable metrics threshold

- 10.4.33 If a subject has some non-reportable core metrics it does not necessarily mean all 6 of its core metrics are non-reportable. It could have reportable metrics for continuation and employment and further study, for example, but have non-reportable NSS metrics.
- 10.4.34 Under this option, we would apply a more lenient definition of 'suitable metrics' that required less than all 6 core metrics to be reportable. We could then apply a different assessment process for different levels of non-reportability.
- 10.4.35 The proposed, more lenient, definition of 'suitable metrics' under this option is that at least two of the following are reportable:
- at least one NSS metric
  - non-continuation metric
  - at least one employment metric, either from DLHE<sup>10</sup> or LEO (selected as a measure of employment outcomes that does not rely on a survey so will be more likely to supply a suitable metric).
- 10.4.36 This option defines different levels of eligibility for assessment, but does not define whether or how assessment would then take place. This option would therefore need to be combined with other options to be implemented into the

---

<sup>10</sup> When TEF is fully implemented in 2019/20 the DLHE will have been replaced by the Graduate Outcomes Survey.

design of subject-level TEF. For example, without all 6 core metrics, an alternative approach to calculating the IH would be needed.

#### Option 2: Group metrics

10.4.37 If assessment goes ahead for a subject with non-reportable core metrics, under this option, assessors would rely on group-level metrics where subject-level metrics are not reportable. Group-level metrics would be produced for each core metric for the 7 subject groups that have been outlined for Model B (these groups are shown in 'Annex A: Subject groupings').

10.4.38 The group-level metrics would be used alongside any reportable subject-level metrics, to ensure that subject-specific metrics still contribute towards the assessment where they exist.

10.4.39 A key question for the **design of this option** is whether providers would be able to choose how the subject groups are defined, or move subjects in or out of the 7 groups, as they are currently permitted to do for submissions under Model B.

- If providers chose their own groups, there would be no consistency across assessments and there may be an incentive for gaming. It also risks undermining the interpretation and robustness of the benchmarks and the operational delivery of this approach may be challenging. The benefit, however, is that the groups may better reflect the way students experience subject provision at that institution and therefore better reflect the performance of the subject being assessed.
- If providers were unable to choose the groups, there would be consistency across providers and it would be easier to produce the metrics, but the group metrics may not align to the way students experience subject provision at all institutions and they may not be as accurate as a proxy for the subject being assessed.

10.4.40 The **benefit** of this option is that the group-level metrics would be a fairly close and sensible proxy, as they include similar subjects to the one being assessed.

10.4.41 The key **disadvantage** of this option is that the group-level metrics would most likely be dominated by the other subjects in the group, meaning the rating could be heavily skewed by these other subjects and there is a risk that the final subject rating could therefore be somewhat misleading to students.

### Option 3: Provider metrics

- 10.4.42 If assessment goes ahead for a subject with non-reportable core metrics, under this option, assessors would rely on the provider-level metrics where subject-level metrics are not reportable.
- 10.4.43 The provider-level metrics would be used alongside any reportable subject-level metrics, to ensure that subject specific metrics still contribute towards the assessment where they exist.
- 10.4.44 The **benefit** of this option is that it is relatively straightforward and does not introduce any new evidence based on metrics into the assessment process.
- 10.4.45 The key **disadvantage** of this option is that provider-level metrics would be a far-removed proxy from the subject in question. The subjects rating would be heavily skewed by other subjects and there is a risk that the final subject rating could therefore be somewhat misleading to students.

### Implementation of Option 2 and 3

- 10.4.46 The implementation of options 2 and 3 would be different across the two proposed models for subject-level TEF.
- 10.4.47 **In Model A**, when identifying exceptions, reliance on provider or group metrics could be applied in two different ways:
- a) The provider- or group-level metrics could be used to ‘fill gaps’ in the subject-level metrics and be used to form a combined initial hypothesis (IH) for the subject. For example, if NSS metrics are non-reportable, the NSS metrics and associated flags from provider or group metrics could be used when calculating the IH for that subject. The resulting subject IH would then be compared to the provider-level IH to identify whether it is an exception. If it were an exception, during the assessment of that subject, the provider- or group-level metrics would be presented alongside the subject-level metrics for assessors to consider as evidence.
  - b) Given that no IH can be generated using the subject-level metrics alone, we could rely entirely on the provider or group metrics to form the IH. For provider-level metrics, we would therefore be assuming that performance of the subject is entirely similar to the provider and it would be treated as a non-exception. This means the subject would either automatically receive the provider-level rating or would receive no rating. For group-level metrics, if there is enough reportable metrics to form an IH, this is used to identify whether the subject is an exception. If not, there would be insufficient evidence that the subject differs from provider-level and it would be treated as a non-exception and receive the provider rating.

10.4.48 **In Model B**, when the subject is assessed, provider or group-level metrics would be presented alongside the subject-level metrics for assessors to consider as evidence.

## 11 Additional evidence

Some sources of additional evidence may be more relevant at subject-level than they were in provider-level TEF. Two key examples are:

- Meeting the standards set out in the QAA Subject Benchmark Statements
- Accreditation or recognition of courses by professional, statutory and regulatory bodies (PSRBs).

While we recognise the importance of this evidence in subject-level TEF, we are not proposing to make the provision of this evidence compulsory. We propose that providers should continue to be free to choose which additional evidence to include and how they present it in their submissions, both at provider and subject-level.

We are consulting on whether there should be mandatory declaration of certain types of evidence, particularly for certain subjects.

### Background

- 11.1 In moving to subject-level TEF, some additional evidence may now be more relevant and important than it was in provider-level TEF. If this additional evidence is more closely linked to the quality of teaching provision at subject level, then it may be appropriate for this evidence to have a greater role in the assessment process for individual subjects.
- 11.2 There are two key examples of additional evidence that may fall into this category:
- Meeting the standards set out in the QAA Subject Benchmark Statements
  - Accreditation or recognition of courses by professional, statutory and regulatory bodies (PSRBs).
- 11.3 There may be other sources of evidence that are not set out here but are also highly relevant to subject-level assessments. In responding to the consultation, we welcome further suggestions from stakeholders about additional evidence that could be particularly important for assessors to consider when undertaking subject assessments, either for all subjects, or for specific subjects. This could be qualitative information or additional metrics that are subject-specific.

### QAA Subject Benchmark Statements

- 11.4 Subject Benchmark Statements set out expectations about standards of degrees in a range of subject areas. They describe what gives a discipline its coherence

and identity, and define what can be expected of a graduate in terms of the abilities and skills needed to develop understanding or competence in the subject.<sup>11</sup> These have been developed in consultation with the higher education sector and are available on the QAA website.

- 11.5 If providers are able to produce evidence that they are meeting the standards and principles set out in these subject statements, this should indicate that their provision of that subject meets the quality benchmark set by the sector for that subject. This would be relevant for assessors to consider when undertaking subject assessments.

### **PSRB accreditation**

- 11.6 Accreditation is a process for verifying or approving a higher education programme or higher education provider. Accreditation is usually granted by PSRBs, which are a diverse group of professional and employer bodies, regulators and those with statutory authority over a profession or group of professionals. PSRBs engage with higher education as regulators, accrediting or endorsing courses that meet professional standards, provide a route into a profession or are recognised by employers.
- 11.7 Accreditation of a course may indicate that the course results in membership, chartered status or full or partial exemption from exams. Accreditation may also signal that the course is recognised as meeting a particular set of criteria or a quality threshold that is recognised by employers.<sup>12</sup> In addition to accreditation, PSRBs may also offer membership or other forms of recognition for a course or provider.
- 11.8 Where there is evidence that PSRB accreditation or recognition improves or provides evidence of the quality of teaching provision and/or student outcomes, this is clearly relevant to the TEF assessment. For this reason, the TEF Specification already includes 'recognition of courses by PSRBs' as an example of possible types of additional evidence for inclusion in provider submissions. As PSRBs are generally based around a discipline or industry, accreditation is particularly relevant to subject-level assessments.

### **Proposal and rationale**

- 11.9 We believe the written submissions are the best way for additional evidence to be considered in the provider and subject assessments in subject-level TEF. We

---

<sup>11</sup> This description is based on information provided by the Quality Assurance Agency (QAA) on their website. For more information, please see <http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements>

<sup>12</sup> This description is based on information provided by the Higher Education Statistics Authority (HESA) on their website. For more information, including a list of PSRBs and HESA's accreditation information table, please see: [https://www.hesa.ac.uk/collection/c16061/accreditation\\_guidance](https://www.hesa.ac.uk/collection/c16061/accreditation_guidance).

therefore do not propose to be prescriptive about the presentation or use of any additional evidence in subject-level TEF. The inclusion of additional evidence in the submission would remain voluntary, with no current proposal to introduce mandatory requirements for any additional evidence beyond those already required for the metrics.

- 11.10 This includes evidence of meeting the QAA Subject Benchmark Statements as well as the role of Professional, Regulatory and Statutory Body (PSRB) accreditation. Providers may choose to cite evidence of these in their subject submissions (and the provider submission if relevant) if they consider it helpful in demonstrating one or more of the TEF criteria. However, this would not be a mandatory requirement imposed on providers, and providers would not be penalised for choosing not to include this information.
- 11.11 The current TEF Specification includes a list of possible examples of evidence for each aspect of quality under the TEF. We recognise that individual subjects may have subject-specific evidence that providers may wish to present in their subject submissions. When developing the list of possible examples for subject-level TEF, we will work with subject bodies to ensure it includes sources of evidence that may be particularly relevant to individual subjects.

### **QAA Subject Benchmark Statements**

- 11.12 The subject benchmark statements are qualitative in nature and set out a series of standards and principles. To identify whether these statements had been sufficiently met, judgement would be required to interpret the provider's approach to interpreting and applying them.
- 11.13 The statements are intended to 'allow for flexibility and innovation in programme design within an overall conceptual framework established by an academic subject community'.<sup>13</sup> Given this, providers may apply these statements in very different ways. It would therefore be challenging to establish a consistent way in which we would expect providers to report against these statements for the purpose of TEF assessment.

### **PSRB accreditation**

- 11.14 The nature of PSRBs, of which there are a large number, varies significantly across the higher education sector. Some of these bodies have a close relationship with a Regulator or Government Department, whereas others are entirely industry-led. Since there is no central body that endorses or regulates PSRBs, the standards set by their accreditation schemes may vary significantly.

---

<sup>13</sup> Quality Assurance Agency, information provided on their website at <http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements>.

There are also differences in the level of coverage and recognition that PSRB accreditation has within different disciplines and industries.

- 11.15 Given this variability, the value and impact that PSRB accreditation has for students may not be equal across all PSRBs. The value for students may also vary by subject depending on the role that accreditation plays in qualification requirements for students planning to work in a field relevant to the accreditation. Accreditation is more important for a graduate's employment prospects in some subjects compared to others.
- 11.16 As it may not be relevant to all subjects, and may apply in different ways, it seems appropriate to allow providers to make their own judgement on how relevant this evidence is and if relevant, how they wish to present it.

### Alternatives

- 11.17 The alternative approach would be to require providers to make a mandatory declaration to cover this additional evidence. This could apply to all or just some subjects.
- 11.18 As outlined above, we do not see how a consistent mandatory declaration could be applied effectively to QAA Subject Benchmark Statements. Similarly, we do not think it is appropriate to mandate a declaration about accreditation across all subjects.
- 11.19 **We do acknowledge however, that there may be particular subjects in which PSRB accreditation should play a greater role in the assessment.** This may be the case for highly regulated subjects where accreditation forms an essential element of a student's progression to employment. If accreditation is a formal or regulated requirement to be considered for employment in that field, then this could have a significant impact on student outcomes and should be considered in the assessment process.
- 11.20 **It may therefore be worth considering whether to introduce a mandatory declaration about PSRB accreditation for certain subjects.**
- 11.21 This approach would introduce the concept of having a slightly different assessment process for each subject, as this additional evidence would only be a formal part of the assessment process for some subjects. As this could introduce some inconsistency across subject assessments in TEF, there would need to be a strong rationale for doing so.
- 11.22 This approach would also increase the burden imposed by the TEF for providers, as an additional declaration would demand further preparation. If the rationale for introducing this approach relates only to the employment prospects of students, it

could be argued that the existing metrics about employment already capture this effect.

**11.23 To fully assess this alternative, we are seeking feedback on this from the sector, including evidence that demonstrates why this approach may be important for certain subjects.**

11.24 In considering the design of this alternative, we would also need to consider how it might affect the incentives of PSRBs. We would want to avoid any incentive for PSRBs to lower their standards or focus on evidence provision for TEF, in any way that risks compromising the focus on meeting the needs of students, professionals and employers.

## 12 Interdisciplinarity

In this chapter we discuss how subject-level TEF will be applied to interdisciplinary provision. We are proposing specific approaches for joint versus multi-subject programmes.

- For **joint programmes**, the two subjects which make up a course will be treated the same as their equivalent single subject programmes. A provider will not be given a separate rating for its joint programmes. Students looking to study a joint honours programme will be able to look at the ratings for each component subject.
- To capture **multi-subject (or combined) programmes**, we are proposing to use three broad 'general' subjects for assessment and ratings. Students looking to study a multi-subject degree will be able to look at the ratings for these general subjects to understand the teaching excellence and student outcomes of these courses.

### Background

- 12.1 The Department for Education recognises that interdisciplinary provision can have a number of benefits, including cross-fertilisation of practice and concepts. The British Academy's recent call for evidence on interdisciplinarity found it to have an "essential role in addressing complex problems and research questions posed by global social challenges" and cited "the increased rigour it can bring to one's understanding of one's own discipline".<sup>14</sup> Real world problems may not fit into subject or discipline classifications, and when knowledge is applied in the wider world, it often requires multi-dimensional input.
- 12.2 Under subject-level TEF, students and courses will be assigned to the 35 CAH2 subject groups and these will inform subject-level metrics and submissions. We know that provision at many providers will cross the boundaries of any subject or discipline definitions we use. This may be particularly challenging for providers for whom interdisciplinarity is an important part of their provision, whether that is through embedding interdisciplinary thinking in their single subject curricula, offering a wide range of joint honours programmes, encouraging students to take up modules outside of their subject area, or through other approaches.
- 12.3 This section outlines how we propose to account for interdisciplinarity in subject-level TEF. We are seeking feedback on the proposed approach to inform the development of any further solutions or adjustments to ensure that, as an

---

<sup>14</sup> Please see <https://www.britac.ac.uk/sites/default/files/Crossing%20Paths%20-%20Full%20Report.pdf>.

unintended consequence, providers are not discouraged from taking an interdisciplinary approach in their provision should this be their preferred strategy.

## Proposal and rationale

### Joint programmes

- 12.4 In the specific case of joint programmes, the two subjects which make up a course will each be treated the same as their equivalent single subject programmes (see Table 4 for an example).
- 12.5 Where the interdisciplinary nature of the joint programme improves the quality of teaching and leads to better student outcomes, this will be relevant in addressing a number of the TEF criteria and may also show up in better metrics performance for each of the individual components versus the equivalent single programmes.
- 12.6 **To achieve this**, students will be counted pro rata against each subject **in the subject-level metrics**. The pro rata proportions used to make this allocation are specific to each course and are based on data provided by individual providers. When making HESA data returns, providers include information on the pro rata allocation to subjects that should be used for each of their courses. An example of this is shown in Table 4.

**Table 4: Example of pro rata proportions for joint courses**

Course	Subject	Pro rata proportion
English literature and Drama Studies	English Studies	0.5
	Creative arts and design	0.5
Psychology and Sociology	Psychology	0.65
	Sociology, social policy and anthropology	0.35

- 12.7 **For submissions**, this proposed approach means that providers may not have the opportunity to address an entire joint programme in a single submission. However, where submissions form part of the assessment, a provider can discuss the joint programme in the submission for each of the individual subjects that comprise it.
- 12.8 **The provider will not be given a separate rating for its joint programme.** It would seem reasonable to expect that a student seeking to study, for example, “Engineering and French”, would be able to look to the ‘Engineering’ and ‘Languages, linguistics and classics’ ratings in order to understand the teaching excellence picture for their joint programme. This approach is being tested with

students through a research project that has been commissioned by the Department for Education.

### Multi-subject and combined programmes

- 12.9 This section refers to programmes and courses that substantially cover more than 2 of the 35 subject groups in CAH2. We have referred to these as multi-subject or combined programmes.
- 12.10 Multi-subject programmes are likely to pose a greater challenge for subject-level TEF than joint programmes, as students may be enrolled in a broad-based degree, possibly choosing their subject after their first one or two years.
- 12.11 Although providers across the UK offer multi-subject degrees, input from Universities Scotland and providers in Scotland has suggested that this type of provision is particularly prevalent in Scotland, and that it is not possible to map at least the first year of many degrees in Scotland against a particular subject.
- 12.12 In addition to these multi-subject degrees, there are also other combined degrees where no subjects can be mapped to the course, or there are more than 3 relevant subjects and providers are unable to provide pro rata allocations in their HESA data returns. In this case, students on such courses are currently mapped to a "combined" subject.
- 12.13 The CAH2 includes three subjects which should accommodate a large proportion of multi-subject and combined provision:
- General and others in sciences<sup>15</sup>
  - Humanities and liberal arts
  - Combined and general studies.
- 12.14 Analysis using data from providers with suitable metrics who entered TEF in academic year 2016/17 indicated that these categories would represent approximately 2% of all instances of subjects within providers across the sector.
- 12.15 **We are proposing that these three general subjects be used to capture provision of multi-subject and combined programmes.** These subjects would be assessed and receive subject ratings in the same way as other CAH2 subjects. Providers offering multi-subject or combined degrees will receive subject-level metrics for these subjects and where relevant, would write an associated submission.

---

<sup>15</sup> We note that there are currently no students mapped to 'General and others in science'. It appears that this is mostly due to the way in which JACS codes have been mapped across onto CAH2. We are assuming that once CAH is fully implemented in the sector, this will change and modular degrees related to science will be mapped to this general subject group.

- 12.16 This approach means that many providers will have the opportunity to address an entire multi-subject or combined programme in a single submission, as multi-subject courses will likely map to only one of these three general subjects. The way in which courses are mapped to subjects will continue to be determined by providers. Where a course covers 3 subjects in the current HESA returns, providers could either continue to map that course to three subjects, or could choose to map that course to one of the three more general subject categories.
- 12.17 While providers will not be given a separate rating for each of its multi-subject or combined programmes, the three general subjects will reflect this specific type of provision. Students looking to study a multi-subject or combined degree will be able to look at the ratings for these general subjects to understand the teaching excellence and student outcomes of these courses. The provider-level rating may also be more relevant to these students, given they may be exposed to several departments across the provider.
- 12.18 We are seeking views from providers who offer multi-subject and combined degrees in order to better understand whether this proposed approach will work. We also encourage any providers that have subject-level metrics for the three general CAH2 subjects listed above to comment on this proposal.

### **Other interdisciplinary approaches**

- 12.19 The TEF seeks to recognise excellence wherever it is found, be it in interdisciplinary or single-discipline teaching. The TEF criteria do not require providers to demonstrate an interdisciplinary approach in making the case for excellence, since TEF takes a broad definition of teaching excellence and does not seek to prescribe a single approach to high quality teaching and good student outcomes. However, since TEF is heavily focused on outcome measures, providers that choose to pursue an interdisciplinary approach should be able to demonstrate whether and how this approach leads to positive outcomes for their students, and therefore why they have chosen to pursue it.
- 12.20 In some subjects, the advantages of an interdisciplinary approach may be reflected in metrics data and require no further explanation. Where providers wish to draw out the impact of their interdisciplinary approach separately from the TEF metrics, they will be able to do so through their written submissions. Furthermore, both models being proposed will incorporate a submission element that will inform the overall provider assessment, in which interdisciplinary approaches can be evidenced, if appropriate.

## 13 Teaching intensity

As part of subject-level TEF, we are also consulting on whether to introduce a new measure of teaching intensity. The measures could be used as part of the TEF assessment process, or presented as stand-alone information for students. If used as part of the TEF assessment, they would only apply to subject-level assessments and would not be used to compare teaching intensity between different subjects.

The Government is aware that factors such as contact hours matter to students' perceptions of their studies. The amount and quality of teaching time that students receive is one important factor that affects their learning and education. It is also a factor that is directly under the control of the provider offering the teaching. The Government considers that excellent teaching is likely to demand a sufficient level of teaching intensity in order to provide a high quality experience for the student.

We recognise that teaching and learning takes place in different ways and understand the complexities around capturing these in a single measure. We are therefore seeking views on introducing a measure of teaching intensity and the positive impacts or unintended consequences of implementing this. This chapter presents several approaches on how teaching intensity could be measured and we welcome feedback on these suggestions.

### 13.1 Introduction

13.1.1 As part of subject-level TEF, we are also consulting on whether to introduce a new measure of teaching intensity. The measures could be used as part of the TEF assessment process, or presented as stand-alone information for students. If used as part of the TEF assessment, a measure of teaching intensity would only apply at subject-level, not at provider level, and measures would not be used to compare teaching intensity between different subjects. The chapter sets out options on how teaching intensity could be measured and used as part of subject-level TEF.

13.1.2 The amount and quality of teaching time that students receive is one important factor that affects their learning and education. It is also a factor that is directly under the control of the provider offering the teaching to them. The Government is aware that factors such as contact hours matter to students' perceptions of their studies. The Higher Education Policy Institute and Higher Education Academy (HEPI-HEA) Student Experience Survey<sup>16</sup> has repeatedly shown that students have falling perceptions of value for money. Compared with the 2016

---

<sup>16</sup> Respondents to the survey are full-time undergraduate students.

survey, there was a 2% decline in those who feel they have received value (good or very good) and a 2% increase in those who feel they have received poor value resulting in almost as many students (34%) who felt they had received poor value as good value (35%) in 2017<sup>17</sup>. Palfreyman<sup>18</sup> believes that if students had certainty over the quantity of teaching and assessment then much dissatisfaction with value for money could be addressed. The Government considers that excellent teaching is likely to demand a sufficient level of teaching intensity in order to provide a high quality experience for the student.

13.1.3 We understand; however, that teaching intensity is difficult to measure, particularly given the rich diversity of pedagogy and the difference made by both varying class sizes and the efficacy of the teaching. Teaching intensity is a measure that goes beyond just counting contact hours, it gives a more rounded picture of the nature and amount of teaching received.

13.1.4 We are therefore seeking views through the consultation on options for measuring teaching intensity. Alongside this, we are piloting two of these options (option 1 and 2, see page 71) as set out in the Subject level pilot specification<sup>19</sup>.

13.1.5 This chapter is structured in the following two sections: rationale for measuring teaching intensity and options for measuring teaching intensity.

## 13.2 Rationale for measuring teaching intensity

13.2.1 As set out by the previous Government in the White Paper: *Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice* (May 2016)<sup>20</sup>, there is strong evidence that teaching intensity, including factors such as contact hours and class size, matter to students' perceptions of their studies. The amount of teaching that a student receives can vary across providers, for example, The Which University Guide gives a range of 13 to 21 hours per week as the teaching time for a full-time mathematics course<sup>21</sup>. Huxley and Peacey<sup>22</sup> also found that when comparing the same course at different providers, the amount of teaching that a student receives can vary. Furthermore,

---

<sup>17</sup> <http://www.hepi.ac.uk/wp-content/uploads/2017/06/2017-Student-Academic-Experience-Survey-Final-Report.pdf>, page 12.

<sup>18</sup> Palfreyman, David. 'The TEF by 2020?' OxCHEPS Occasional Paper. Oxford, UK: Oxford Centre for Higher Education Policy Studies, July 2016.  
[http://oxcheps.new.ox.ac.uk/MainSite%20pages/Resources/OxCHEPS\\_OP60.pdf](http://oxcheps.new.ox.ac.uk/MainSite%20pages/Resources/OxCHEPS_OP60.pdf).

<sup>19</sup> <https://www.gov.uk/government/publications/teaching-excellence-framework-subject-level-pilot-specification>

<sup>20</sup> <https://www.gov.uk/government/publications/higher-education-success-as-a-knowledge-economy-white-paper>

<sup>21</sup> <https://university.which.co.uk/subjects/math>

<sup>22</sup> Class size at university, page 1 (<https://doi.org/10.1111/j.1475-5890.2017.12149>)

Gibbs' dimensions of quality<sup>23</sup> shows that large class sizes can have negative impacts on access to teaching staff, assessment and feedback, student engagement and depth of learning.

- 13.2.2 The 2017 HEPI-HEA Student Experience<sup>24</sup> survey shows that students value contact hours. It reports the percentage of students with different levels of contact hours who are satisfied with the hours they have (see chart on page 31 of the report). Levels of satisfaction were the lowest amongst students receiving 0-9 hours with 55% of students satisfied in 2017. The peak level of satisfaction with contact hours was between 10 and 19 hours with 71% of students satisfied (with 69% satisfied with 20 to 29 hours and 65% satisfied with 30+ hours).
- 13.2.3 It is clear that providers, by their actions, recognise that the amount and type of teaching provided is a critical component of a student's education. Through their decisions on how much, what form and what staff-student ratio of teaching to allocate to a given course, a provider is making an explicit declaration of what teaching they consider to be best for their students, within the relevant resource constraints. Furthermore, all these factors – alongside others, such as course design, resources and student support – are likely to be considered by a provider when considering how to improve the teaching or outcomes of a course.
- 13.2.4 As teaching represents a direct and significant cost to a provider, it follows that, if a provider genuinely believed it was possible to provide an equally high quality academic experience with less teaching intensity, it would be able to reduce the level of teaching and either invest the money elsewhere, or alternatively reduce fees. In fact, and in contrast to this, providers that are seeking to provide a high quality teaching experience frequently seek to invest additional resources to ensure that students receive higher levels of teaching intensity.
- 13.2.5 Measures of teaching intensity would support the Teaching Quality aspect of the Teaching Excellence and Student Outcomes Framework (TEF) by improving measurement of Student Engagement. The criterion states “teaching that provides an appropriate level of contact, stimulation and challenge, and which encourages student engagement and effort.”<sup>25</sup>
- 13.2.6 A measure of teaching intensity could be considered as part of the TEF assessment. The Government considers that excellent teaching is likely to demand a sufficient level of teaching intensity in order to provide a high quality experience for the student. Providers should therefore be investing resources

---

<sup>23</sup> Gibbs, G., 2010. *Dimensions of quality*. York: Higher Education Academy.

<sup>24</sup> <http://www.hepi.ac.uk/wp-content/uploads/2017/06/2017-Student-Academic-Experience-Survey-Final-Report.pdf>

<sup>25</sup> Teaching Excellence and Student Outcomes Framework Specification.

<https://www.gov.uk/government/publications/teaching-excellence-and-student-outcomes-framework-specification>, p24

into teaching, measurable through the volume of contact time provided, the sizes of the classes in which teaching is delivered, or a combination of the two.

## Approaches to teaching

- 13.2.7 It is an important principle when considering teaching intensity that the Government accepts the right of providers to decide how teaching should be carried out and is not beginning with a view on whether certain types of teaching methods are better than others. The higher education sector comprises a wide range of philosophies and approaches to teaching, and this diversity is a key feature of the sector. The Government recognises that activities not captured in a measure of teaching intensity may also be important to a student. Any teaching intensity measure should encompass what is most relevant to students.
- 13.2.8 Teaching and learning activities that would be included in any teaching intensity measure are defined broadly using the QAA taxonomy<sup>26</sup> as a starting point. Examples of some of the factors involved in measuring teaching intensity include size of the teaching group, where the teaching is taking place (e.g. face-to-face in a classroom, or online) and structure of the activity (e.g. lecture or demonstration).
- 13.2.9 Options for measures of teaching intensity vary from simply measuring the amount of contact time only, to also capturing the level of interaction (e.g. staff student ratio), external placements and online teaching.
- 13.2.10 One of the areas on which there appears to be less consensus is whether or not independent study should be included in a measure of teaching intensity. A critical part of higher education is the ability to learn how to study independently; furthermore, for any given student, the amount of independent study undertaken is likely to correlate, other factors being equal, to the amount learned. The drawbacks, however, of including independent learning as part of a measure of teaching intensity are that it does not actually measure what teaching a student is receiving (and hence does not measure value for money) and is more dependent on the student than on the provider. Furthermore, it is difficult to reliably collect data on students' independent study.

---

<sup>26</sup> *Explaining Contact Hours: Guidance for institutions providing public information about higher education in the UK*, August 2011, <http://www.qaa.ac.uk/en/Publications/Documents/contact-hours.pdf>. QAA defines teaching activities to include: lecture; seminar; tutorial; project; demonstration; practical classes and workshops; supervised time in studio/workshop; live online sessions; one-to-one staff time (eg office hour).

## 13.3 Measures of teaching intensity

13.3.1 This section presents and describes various options on how information on teaching intensity could be captured and discusses some advantages and disadvantages of each.

13.3.2 The options presented are:

1. Gross Teaching Quotient (GTQ), external visits and work-based learning, and online teaching
2. Student survey on contact hours
3. GTQ weighted by qualification/seniority of teacher
4. A measure using quantitative and qualitative information about how a student is expected to spend their time on a course
5. Measure of engagement with teaching resources
6. Measure of staff contracted teaching hours

13.3.3 Methodologies around the first option have been developed by Huxley *et al*<sup>27</sup> and the second option is run by HEA and HEPI<sup>28</sup>. Variations on these themes have been further developed by the Department for Education and HEFCE and are being tested in the subject level TEF pilot. See the [HEFCE guidance](#) document for more information on the details of these measures.

13.3.4 Currently data on many of the options for measuring teaching intensity (see Table 5 below) is not routinely collected by providers so introducing a measure would require new data to be collected. Recognising the diversity of the sector, such data collection should be sufficiently flexible to not require the adoption of specific software or processes by all providers. Furthermore, the Government considers it important that data collection in this area should not itself drive teaching practices, nor impinge institutional autonomy by mandating activities that a provider may consider unfavourable to students or contradictory to its ethos of teaching, such as mandatory attendance monitoring. We would like to get a view on the burden of data collection for any teaching intensity measure and explore the feasibility of using existing collections whilst teaching intensity measures are explored.

---

<sup>27</sup> Huxley, G., Mayo, J., Peacey, M. W. and Richardson, M. Class Size at University. Fiscal Studies. Accepted Author Manuscript. doi:10.1111/j.1475-5890.2017.12149

<sup>28</sup> <http://www.hepi.ac.uk/2017/06/07/2017-student-academic-experience-survey/>

13.3.5 Across any measure of teaching intensity, there are general principles that would need to be taken into consideration. For example, measures should be applicable across different subjects and modes of delivery. Some of these considerations will be explored in the pilots. The higher education sector is diverse offering a huge range of teaching provision including offering courses on a full-time basis, part-time basis, sandwich courses, accelerated courses (a full-time course that is completed in less time than a standard full-time course), single/joint honours courses, Higher National Diplomas (HND) and other sub-degree provision and top-ups from sub-degree to honours degree. This is a strength of the sector and any teaching intensity measure should capture this diversity. Teaching intensity measures would be reported at subject level and therefore will need to factor in joint honours programmes and the aggregation process.

## Summary of options for measuring teaching intensity

**Table 5: Summary of options for measuring teaching intensity**

Option	Description	Comment
1. Gross Teaching Quotient (GTQ); external visits and work-based learning; and online teaching	The GTQ measures not only teaching time but also class size. External visits and work-based learning reflects scheduled learning activity that occurs outside of usual face-to-face teaching. Online teaching is the amount of time spent facilitating online learning.	<ul style="list-style-type: none"> <li>Measures all scheduled teaching time, recognising that this does not always happen face-to-face.</li> <li>Does not take account of independent learning or quality of those doing the teaching.</li> </ul>
2. Student survey on contact hours	Measure based on asking students about the scheduled teaching received (quantity and perception) and their own personal experience of the teaching.	<ul style="list-style-type: none"> <li>Ensures students' views are accounted for.</li> <li>Perception might not reflect actual provision.</li> </ul>
3. GTQ weighted by qualification/ seniority of teacher	Instead of weighting by class sizes, as in option 1, GTQ would also weight contact time by qualification/ seniority of the teacher. The qualification and seniority of the teacher could be seen as proxies for the quality of the teaching.	<ul style="list-style-type: none"> <li>Qualification of the teacher is relevant to the student experience of contact time.</li> <li>No consensus on what would be a good proxy for 'good teacher' (e.g. teaching qualification, PhD, research active, years of experience in industry).</li> </ul>

Option	Description	Comment
4. A measure using quantitative and qualitative information about how a student is expected to spend their time on a course	Measure based on providers stating information on how, for example, each 20 credit module in a course is experienced as learning by students – including both taught hours and expectations of independent learning. This type of data used to be collected as part of the Key Information Sets.	<ul style="list-style-type: none"> <li>• Similar to the Key Information Sets (KIS) but aggregated to subject-level.</li> <li>• Variation in module organisation across providers makes it difficult to compare providers.</li> <li>• Independent learning declarations not readily auditable.</li> </ul>
5. A measure of engagement with teaching resources	Measure of engagement with teaching resources using data from a greater range of inputs including use of libraries and digital resources, completion of assignments and other matters.	<ul style="list-style-type: none"> <li>• Takes into account different patterns of study.</li> <li>• Data collection may be very intrusive.</li> <li>• Not clear how one would combine the inputs into a usable measure valid across different types of provision.</li> </ul>
6. Measure of staff contracted teaching hours	Measure using the number of hours' staff are contracted to teach, and compare to number of students.	<ul style="list-style-type: none"> <li>• In theory, it shows exactly how many hours are devoted to teaching per student.</li> <li>• Highly aggregated and excludes independent study or quality of teacher.</li> </ul>

### Option 1: Gross Teaching Quotient (GTQ) plus external and online teaching

13.3.6 The first option proposes a definition of teaching intensity to measure not only teaching time but also class size based on the method proposed by Huxley *et al*<sup>27</sup>. The proposal would be to capture information in the following categories:

- Gross Teaching Quotient (GTQ)
- External visits and work-based learning
- Online teaching.

13.3.7 The GTQ is calculated by multiplying the taught hours by the appropriate weighting and summing across all weighting bands. The weighting bands are based on the staff/student ratio and are chosen to broadly reflect different class sizes (small group sessions, lectures, seminars etc.). See 'Annex E: GTQ example' for a worked example of how this is calculated.

13.3.8 'External visits and work-based learning' reflects scheduled learning activity in taught years of study that occurs outside of usual face-to-face teaching, which

typically involves unsupervised activity or activity supervised by staff or appointed representatives. This would be an optional category; some courses might have no external visits or work-based learning.

13.3.9 Online teaching reflects the amount of time staff spend in facilitating online learning when they are not necessarily online at the same time as the students. As this sort of online teaching is usually not scheduled, it cannot easily be returned with the scheduled teaching activities in the GTQ. This would be an optional category; some courses might have no online teaching component.

### Advantages

13.3.10 The GTQ is fundamentally an input measure, so providers would have the choice of whether to offer a greater number of small group sessions, or a lesser amount of large group sessions. The GTQ model values 10 hours spent in a group of 30 and 5 hours spent in a group of 15 at the same level (assuming both are taught to the same number of students by the same number of staff). If students have different preferences in terms of how they like to learn, then providers could use this information to differentiate themselves in the types of provision they offer.

13.3.11 The GTQ measures scheduled teaching activity that is provided directly by members of staff in real time, either face-to-face or online.

13.3.12 This option recognises that not all teaching and learning activity is provided face-to-face on campus. The categories 'external visits and work-based learning' and 'online teaching' are included in order to help contextualise why a GTQ may be low, for instance for a distance-learning course taught online or a course including a high number of work placements. It is also recognised that some courses may have no provision in the 'external visits and work-based learning' and 'online teaching' categories.

### Disadvantages

13.3.13 It is difficult to know what value of GTQ should be aimed for. A high value for GTQ may be good value for money but it does not necessarily indicate better learning.

13.3.14 It does not take into account independent learning undertaken by students. The emphasis is on providing organised activities. This option may encourage providers simply to increase organised learning sessions at the expense of encouraging students to become independent learners.

13.3.15 It is vulnerable to gaming as providers could use less experienced staff (for example, PhD students) to increase the number of small group teaching

sessions. The use of weighting bands also creates a risk of gaming around the band boundaries.

## **Option 2: Student survey on contact hours**

- 13.3.16 Option two proposes a measure based on students' perception of contact time using a student survey on contact hours. The student survey would be completed by undergraduate students in each year of study. The survey is based around the scheduled teaching received by a student and their own personal experience of the teaching.
- 13.3.17 The student survey would ask students to reflect on the teaching they have received in a term, week or year, and consists of questions around scheduled face to face teaching, tutor-led online learning activities, off-site teaching (such as fieldwork and external visits) and independent learning.
- 13.3.18 The survey could ask questions such as: 'how many hours per week were scheduled for face-to-face teaching such as lectures, seminars, tutorials, project supervision, demonstrations, practical classes & workshops, supervised time in a studio/workshop etc?'; 'how many hours per week on average did you spend in tutor-led online learning activities?' and 'how many hours per week on average did you spend in independent study?'. It could also ask students to reflect on their personal experience of the teaching received asking questions or agreement to statements such as 'there is enough teaching (face-to-face or online) to support my learning'.
- 13.3.19 This approach is being piloted in the subject-level pilot. To see the full list of questions in the student survey, see the HEFCE guidance document.
- 13.3.20 The timing of any survey based data collection will be an important factor given that students are already expected to participate in a number of other surveys. It may be possible to include questions on teaching intensity as part of the NSS, but a separate survey would still be required for students not covered by the NSS.

### **Advantages**

- 13.3.21 The student survey ensures information about students' views on the adequacy and nature of hours provided is taken into account when measuring teaching intensity. It could also capture information around independent learning time.

### **Disadvantages**

- 13.3.22 This only captures the student's view of the teaching that was provided and/or perceived to be provided. This may not accurately reflect what was on offer from the provider. Questions would need rigorous cognitive testing, as small changes to the wording of questions could have a large impact on responses.

For example, poor phrasing could cause students to respond very differently to a question about their perceptions. In addition, most students only experience one provider and their responses may be based on what they think other providers are offering.

13.3.23 The perception of contact hours may be different for different groups of students. For example, some students may prefer large class sizes and may perceive large classes as positive, whilst other students may prefer small class sizes. However, benchmarking<sup>29</sup> would help to alleviate some of these differences.

13.3.24 Students are already asked to complete a number of other surveys and may suffer from survey fatigue.

### **Option 3: GTQ weighted by qualification/seniority of teacher**

13.3.25 Option 3 proposes a variant to option 1 which would take the GTQ, but instead of weighting by class sizes alone, it would also weight by qualification/seniority of teacher. The reasoning here is that the qualification and/or seniority of the teacher would be proxies for the quality of the teaching. We recognise that in some disciplines practitioner expertise is highly valuable and would need to be recognised somehow if qualification does not account for it. Teaching qualifications could include formal teaching qualifications or non-teaching qualifications such as PhDs.

13.3.26 The Government recognises that teaching in higher education is delivered by a variety of people, including academics, graduate teaching assistants, researchers, expert practitioners from outside of the higher education sector and technicians. The quality of teaching delivered by these individuals will depend on factors that include professional experience and knowledge, teaching-related training, personal attitude, and the provider's investment in high quality teaching.

13.3.27 It is not easy to capture these qualities using only numeric data on teaching hours.

### **Advantages**

13.3.28 The qualification/seniority of the teacher may be relevant to the student experience of contact time. The quality of the teaching may be very different between an experienced lecturer and a PhD student for example. Weighting by such factors would enable greater differentiation between providers and subjects.

---

<sup>29</sup> A process that produces weighted sector averages where weightings are based on the characteristics of the students at the provider.

## Disadvantages

13.3.29 The seniority/qualification of a teacher does not necessarily indicate a better quality of teaching. GTQ weighted by qualification of teacher does not measure the actual quality of the teaching, but would be based on an assumption that certain types of staff are better teachers than other types of staff.

13.3.30 There is no consensus on what a good teacher is. For example, should it be based on having a PhD, a formal teaching qualification, a HEA Fellowship, or a certain number of years experience in industry or professional practice?

### **Option 4: A measure using quantitate and qualitative information about how a student is expected to spend their time on a course**

13.3.31 Option 4 is to ask providers to give information about how students are expected to spend their time on a course.

13.3.32 Providers could be asked to state information on how for example, each 20 credit module<sup>30</sup> in a course is experienced as learning by students. This learning could include face-to-face contact time through lectures/seminars, tutorials and other tasks/preparation work completed in between lectures. The notional learning<sup>31</sup> hours associated with qualifications, programmes and individual units of study are based on a broad agreement across providers that students can expect to spend 10 hours learning on average in order to gain one academic credit unit<sup>32</sup>

13.3.33 Providers would give information on how many hours a student is expected to spend on face-to-face contact (e.g. lectures/seminars), online learning, external learning and preparation work such as completing assignments and independent study.

13.3.34 This information could be obtained through a provider declaration. This type of data was previously collected as part of the Key Information Sets<sup>33</sup> (KIS), which was mandatory for all higher education institutes in England and Wales. The KIS were sets of standardised information about undergraduate courses. They were designed to allow students and other interested parties to compare data between courses and providers. The KIS data is no longer collected but it would be possible to re-instate this data collection.

---

<sup>30</sup> A 20 credit module is used as an example only.

<sup>31</sup> <http://www.qaa.ac.uk/en/Publications/Documents/Academic-credit-in-higher-education-in-England---an-introduction.pdf>, page 6

<sup>32</sup> Credit is a means of quantifying and recognising learning whenever and wherever it is achieved. A number of credits is normally assigned to each module, which indicates the amount of learning undertaken, and a specified credit level indicates the relative depth of learning involved.

<sup>33</sup> <http://www.hefce.ac.uk/lt/unikis/>

## Advantages

13.3.35 Using this approach, a student would be able to get an idea of how they are expected to spend their time for a subject.

13.3.36 The approach captures different types of learning including independent study.

## Disadvantages

13.3.37 This data was difficult to capture on the KIS, partially due to them being prospective not retrospective. There is no easy way of validating whether the declared levels of self-study bear any resemblance to what actually occurs. There would also be strong incentives on providers to exaggerate what is expected.

13.3.38 There may be a lot of variation in how modules are organised (in terms of teaching and learning) across providers and across subjects within providers. This may make it difficult to easily compare providers and aggregate up to subject level.

13.3.39 It is not obvious how different patterns of teaching could be effectively compared.

## Option 5: Measure of engagement with teaching resources

13.3.40 This approach would measure student engagement by using data from a greater range of inputs. This could include lecture/seminar attendance, virtual learning environment usage, library visits and other learning facilities.

13.3.41 This option assumes that use of teaching facilities such as lecture theatres, classrooms, libraries etc. is captured through an electronic data capturing system. For example, if students are required to swipe in with their student identity card every time they use a classroom or the library, this information would be stored.

13.3.42 This data could then be used to create a measure based on which facilities are used and how often.

13.3.43 There would need to be a way of combining all aspects/data to turn this vast information into suitable measures, which would take time to develop and could result in quite complex formulae.

## Advantages

13.3.44 The measure takes into account different patterns of study and enables potential students to gain an insight into the full range of engagement opportunities (with teaching resources) that support academic success in higher education. For example, the measure could identify the range of support offered

by providers, giving students a clear understanding of how they will be supported.

13.3.45 The measure would be based on individual use of teaching resources which would be aggregated up to subject level. It would capture engagement with teaching resources across a broader range of inputs, rather than solely contact hours.

#### Disadvantages

13.3.46 It is not at all obvious how this data could be combined into a small number of indicators that could readily be compared across providers. There is little public evidence base to support any particular combination and any proposed formula would likely be highly contested.

13.3.47 All providers may not necessarily collect data on use of facilities by students, and this measure is reliant on the availability of such data, and the systems needed to capture it. If these are not in place, collecting and validating the data could be too burdensome. Even where the data is collected, it is likely to be difficult to compare in a like-for-like manner across providers.

13.3.48 It is unclear how to combine data from all different aspects into usable indicators. It would require a complex methodology to create a suitable value, which may not be easily understandable.

13.3.49 Different types of students may engage with teaching resources in very different ways. To meaningfully compare different providers, it is likely that a significant amount of benchmarking would be necessary, adding further complexity to the measure.

#### Option 6: Measure of staff contracted teaching hours

13.3.50 Option 6 proposes a measure using the number of hours' staff are contracted to teach. This could include all types of teaching that a staff member is contracted for, including lectures, seminars, tutorials, one-to-one sessions, practical sessions etc.

13.3.51 It might be possible for providers to use their administrative systems to extract the required data, assuming information on contracted hours is held against each member of staff. This could be aggregated to give a total per year for a subject, which would then be compared to the total number of students studying that subject.

#### Advantages

13.3.52 The approach is simple and shows exactly how many hours are devoted to teaching per student.

## Disadvantages

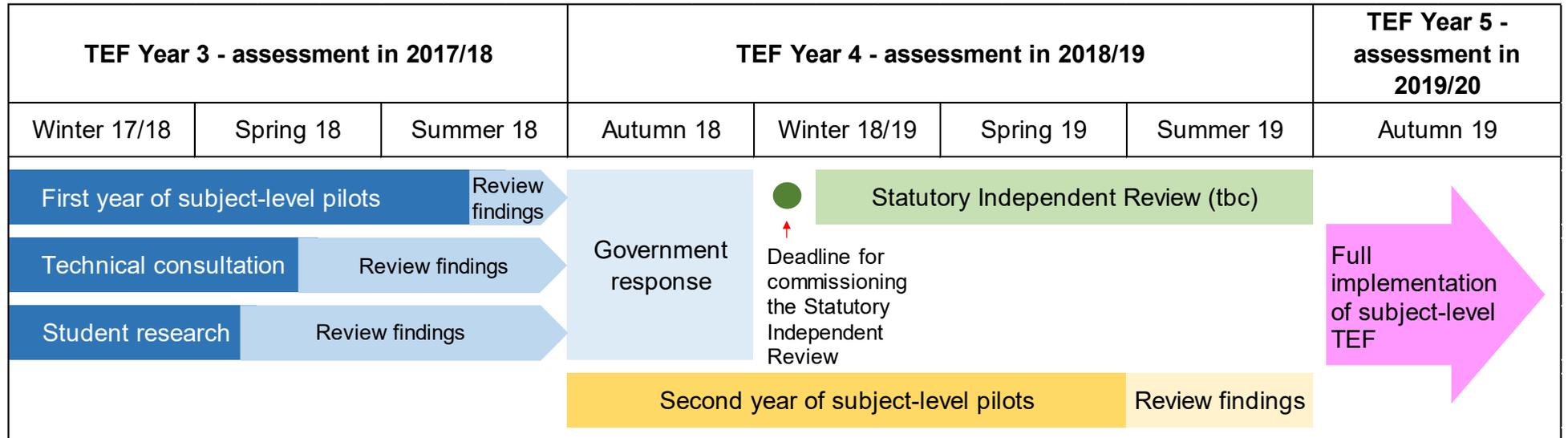
13.3.53 This approach is highly aggregated and does not include any information on the type of teaching or the degree of self-study.

13.3.54 Staff may also be contracted to teach in departments other than their own. For example, if the Maths department teaches Engineering students, should this count towards a teaching intensity measure for the Maths department?

## 14 Next steps and relationship with independent review

- 14.1 The consultation was published on 12 March 2018 and closes on 21 May 2018. During this period, the Department for Education will be holding several events to engage the sector in the consultation.
- 14.2 We encourage stakeholders to offer their views at these events, as well as through the online response form. While the consultation seeks input on specific topics, you are also welcome to comment on any aspect of the detailed design of subject-level TEF, as set out in the [TEF Year Three pilot specification](#).
- 14.3 **We will bring together the findings from all three exercises (the consultation, the first year of subject-level pilots and student research) to form a single Government response document. We expect to publish this in autumn 2018.** This will enable us to give full consideration to the consultation responses within the context of the whole policy development process, once all evidence has been gathered. Evidence from each exercise will be brought together to inform the design of the second year of the subject-level pilots, as well as to help refine the overall design of subject-level TEF as we move towards full implementation, which is planned for academic year 2019/20.
- 14.4 The findings from the consultation will also be available to the Statutory Independent Review of the TEF, which will be conducted during academic year 2018/19. The scope of the review is set out in section 26 of the 'Higher Education and Research Act 2017'.
- 14.5 The timeline for subject-level TEF is shown in Figure 9. As discussed above, the Government response to the consultation will also cover the response to the pilots and student research. The second year of subject-level pilots and the independent review will take place over academic year 2018/19 before full implementation of subject-level TEF, which is expected in academic year 2019/20.

Figure 9: Subject-level TEF timeline



## Annex A: Subject groupings

Table 6 shows the subject classification system proposed for subject-level TEF. **The 35 subjects from the second level of the Common Aggregation Hierarchy will be used** for: metrics, submissions (by exception) and ratings for Model A; and metrics and ratings for Model B. The 7 subject groups will be used for submissions in Model B and could be used for metrics when assessing a subject with non-reportable core metrics.

**Table 6: Subject groups for subject-level TEF**

<b>7 subject groups</b>	<b>CAH2 Codes</b>	<b>35 CAH2 subjects</b>
Medical and health sciences	CAH01-01 CAH02-01 CAH02-02 CAH04-01 CAH02-03 CAH05-01 CAH03-02	1. Medicine & dentistry 2. Nursing 3. Pharmacology, toxicology & pharmacy 4. Psychology 5. Subjects allied to medicine 6. Veterinary science 7. Sport & exercise sciences
Engineering and technology	CAH11-01 CAH10-01 CAH10-02	8. Computing 9. Engineering 10. Technology
Natural sciences	CAH06-01 CAH03-01 CAH07-02 CAH09-01 CAH07-01 CAH07-03 CAH08-01	11. Agriculture, food & related studies 12. Biosciences 13. Chemistry 14. Mathematical sciences 15. Physics & astronomy 16. Physical, material & forensic sciences 17. General & others in sciences
Social sciences	CAH13-01 CAH15-02 CAH12-01 CAH15-03 CAH15-01 CAH22-01 CAH15-04	18. Architecture, building & planning 19. Economics 20. Geographical & environmental studies 21. Politics 22. Sociology, social policy & anthropology 23. Education & teaching 24. Health & social care
Business and law	CAH17-01 CAH16-01	25. Business & management 26. Law
Arts	CAH21-01	27. Creative arts & design

<b>7 subject groups</b>	<b>CAH2 Codes</b>	<b>35 CAH2 subjects</b>
Humanities	CAH19-02 CAH18-01 CAH19-01 CAH19-03 CAH20-01 CAH14-01 CAH20-02 CAH23-01	28. Celtic studies 29. Communications & media studies 30. English studies 31. Languages, linguistics & classics 32. History & archaeology 33. Humanities & liberal arts 34. Philosophy & religious studies 35. Combined & general studies

## Annex B: Absolute value thresholds

The provider-level thresholds for the academic year 2017/18 TEF metrics are shown in Table 7. If a metric indicator is:

- equal to or above the ‘very high’ threshold value shown in the table, it will receive a star (\*)
- equal to or below the ‘very low’ threshold value shown in the table, it will receive an exclamation mark (!).

**Table 7: Very high and low absolute value thresholds for each metric are from TEF data for assessments undertaken in academic year 2017/18**

Metric	Full-time (%)		Part-time (%)	
	Very high	Very low	Very high	Very low
The teaching on my course (NSS)	90.13	80.01	93.11	79.40
Assessment and feedback (NSS)	83.19	66.86	86.30	66.36
Academic support (NSS)	87.30	75.21	89.34	73.07
Continuation	95.78	79.78	88.59	64.98
Employment or further study (DLHE)	97.10	90.58	99.26	94.16
Highly skilled employment or further study (DLHE)	83.01	55.30	90.37	63.88
Sustained employment or further study (LEO)	84.99	77.00	89.06	78.53
Above median earnings threshold or further study (LEO)	77.51	32.43	82.06	57.14

**Notes:** The top and bottom 10 per cent of providers has been calculated (for the metric and mode in question) on the basis of all providers with a reportable metric that refers to more than 100 students.

We are proposing to apply the same thresholds in Table 7 to the subject-level metrics. The alternative approach would be to set subject-specific thresholds. Table 8 shows the maximum and minimum thresholds that would occur across all 35 subjects when the thresholds are subject-specific (i.e. are defined by subject). These are presented alongside the provider-level thresholds for context. This table only includes the thresholds for full-time provision because sample sizes become too small to robustly calculate subject-specific thresholds for all part-time provision.

**Table 8: Very high and low absolute value thresholds across subjects (full-time) for each metric are from TEF data for assessments undertaken in academic year 2017/18**

Metric		Very high (%)	Very low (%)
The teaching on my course (NSS)	As defined at provider-level	90.13	80.01
	Max when defined by subject	95.99	88.34
	Min when defined by subject	87.23	67.63
Assessment and feedback (NSS)	As defined at provider-level	83.19	66.86
	Max when defined by subject	85.30	71.16
	Min when defined by subject	73.84	44.04
Academic support (NSS)	As defined at provider-level	87.30	75.21
	Max when defined by subject	92.23	78.96
	Min when defined by subject	85.27	60.55
Continuation	As defined at provider-level	95.78	79.78
	Max when defined by subject	99.51	96.96
	Min when defined by subject	94.07	74.55
Employment or further study (DLHE)	As defined at provider-level	97.10	90.58
	Max when defined by subject	100.00	99.41
	Min when defined by subject	93.21	81.65
Highly skilled employment or further study (DLHE)	As defined at provider-level	83.01	55.30
	Max when defined by subject	100.00	98.89
	Min when defined by subject	69.60	41.34
Sustained employment or further study (LEO)	As defined at provider-level	84.99	77.00
	Max when defined by subject	90.73	83.62
	Min when defined by subject	82.07	71.33
Above median earnings threshold or further study (LEO)	As defined at provider-level	77.51	32.43
	Max when defined by subject	98.62	90.83
	Min when defined by subject	52.76	26.27

**Notes:** This analysis is based on provider and subject threshold values provided by HEFCE based on TEF metrics data prepared for assessments in academic year 2017/18. There are some cases where the top and bottom 10% would comprise a single provider on account of there being no more than 10 large providers (with >100 students) in that subject for that metric. The thresholds cannot be identified in these cases, and caution should be exercised where the number of large providers is otherwise small.

This is particularly evident for part time metrics, where subject thresholds cannot be calculated at all for the NSS metrics and are extremely limited for the other core metrics and supplementary LEO metrics (i.e. the thresholds can only be calculated for between 3 and 6 subjects). For this reason, we have not reported any part time thresholds in this table.

For the full-time metrics, the minimum and maximum has been calculated excluding certain subjects where this is the case. Specifically, 4 subjects have been excluded for the NSS metrics, 3 for the continuation metric and 4 for the employment/further study and highly skilled employment/further study metrics.

## Annex C: Metrics charts

The box plots below demonstrate the distribution of core metrics for full-time students across all providers, showing how the range of indicators compare for each of the 35 CAH2 subjects (discussed in section 10.3). They are based on TEF data for assessments undertaken in academic year 2016/17. The right hand side shows the number of providers that offer that subject. This ranges from 16 providers for Celtic studies to 319 providers for business and management.

The very high and very low thresholds shown on the graph as the green and blue lines, respectively have been provided by HEFCE based on TEF data prepared for assessments in academic year 2017/18.

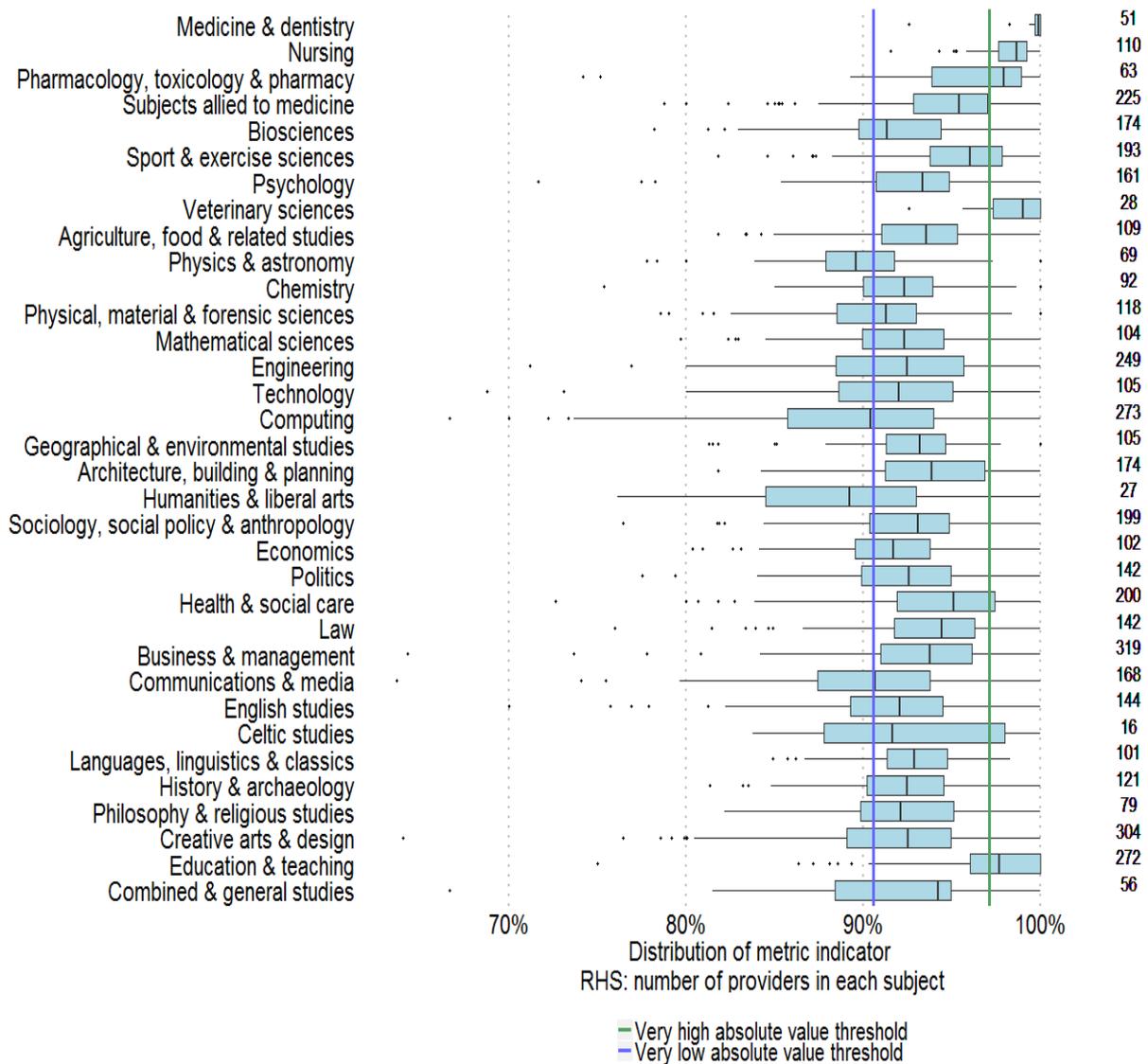
Each boxplot shows the overall pattern of a metric indicator across providers for a subject. The mid-point of the data is shown by the line that divides the box into two parts. Half the indicator values are greater than or equal to this value and half are less. The length of the box or the whiskers indicate how varied the indicator is across providers for a subject and the black dots represent outlying values. For example, using the chart for employment or further study indicators for full-time students, for Celtic studies, there is little variation across providers at the top and bottom (shown by the short whiskers) but there is a lot of variation in the middle (shown by the long box). In contrast for nursing, there is little variation across providers.

The employment or further study metric is the proportion of leavers (responding to the DLHE) who report that they are in employment or further study. Figure 10 shows the distribution of this metric across providers for full-time students by the 35 CAH2 subjects<sup>34</sup>. While median proportions lie consistently between approximately 88% and 100% across subjects, there is significant variation within subjects. For Computing, for example, the difference between the highest and lowest proportion reaches approximately 30 percentage points where as for medicine, it is around 8 percentage points.

---

<sup>34</sup> See note to Figure 10 for coverage of providers.

**Figure 10 Distribution of employment or further study indicators (full-time study)**



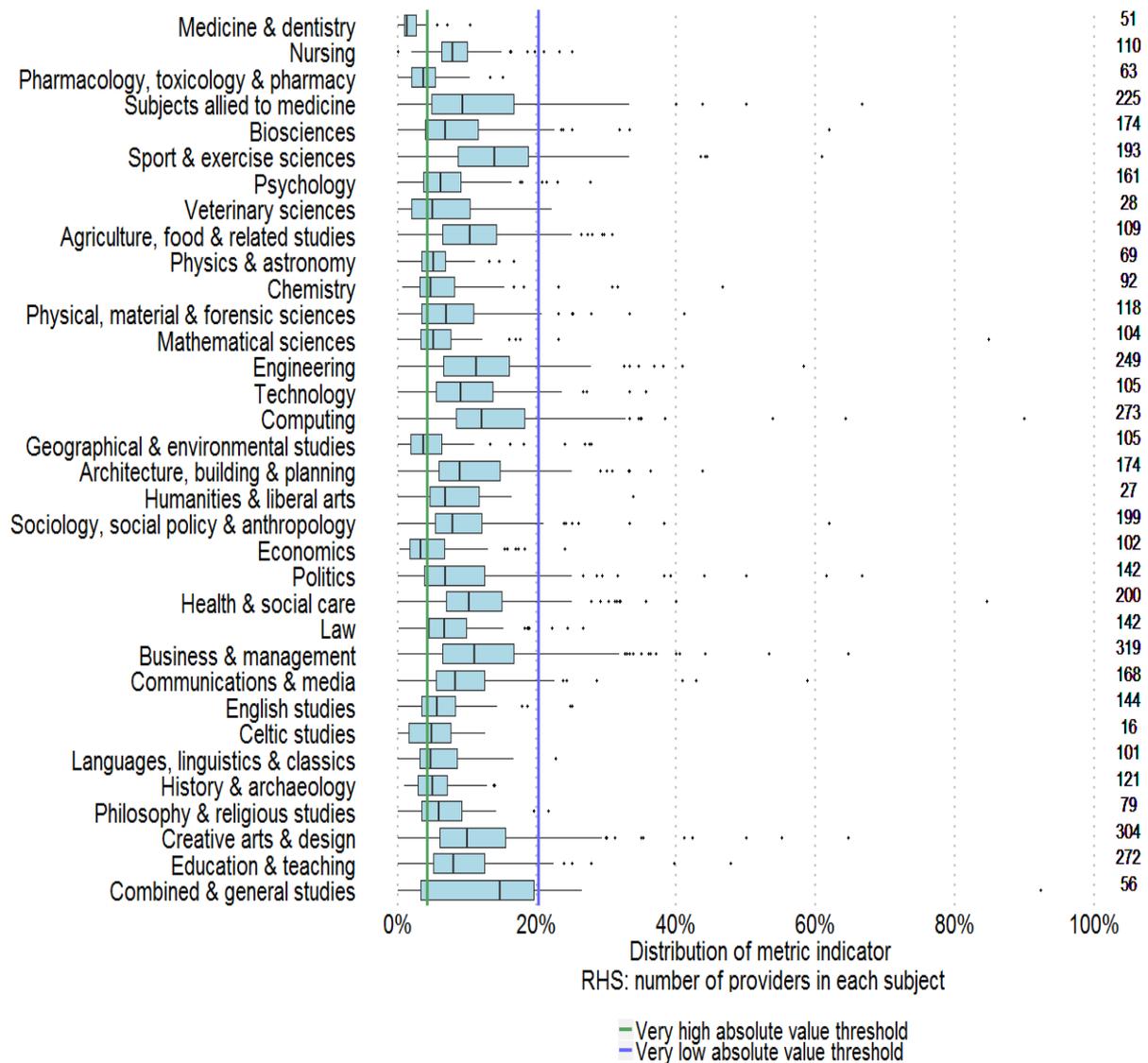
**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low absolute thresholds are from TEF data prepared for assessments in academic year 2017/18.

The non-continuation metric for full-time students tracks students from the year they enter an HE provider to the following year and is presented as the proportion of students who start but do not continue their studies. Figure 11 shows the distribution of this metric across providers for full-time students by the 35 CAH2 subjects<sup>35</sup>. Although the indicators range from 0% to nearly 100% for some subjects, all subjects had indicators in the range 0% to 20% for most providers. A number of subjects showed little variability across providers (eg medicine and dentistry; and physics and astronomy). It

<sup>35</sup> See note to Figure 11 for coverage of providers.

should be noted that following the TEF Year 2 lessons learned exercise, the non-continuation metric will be presented as 'continuation', so that all metrics point in the same direction (i.e. higher values are positive).

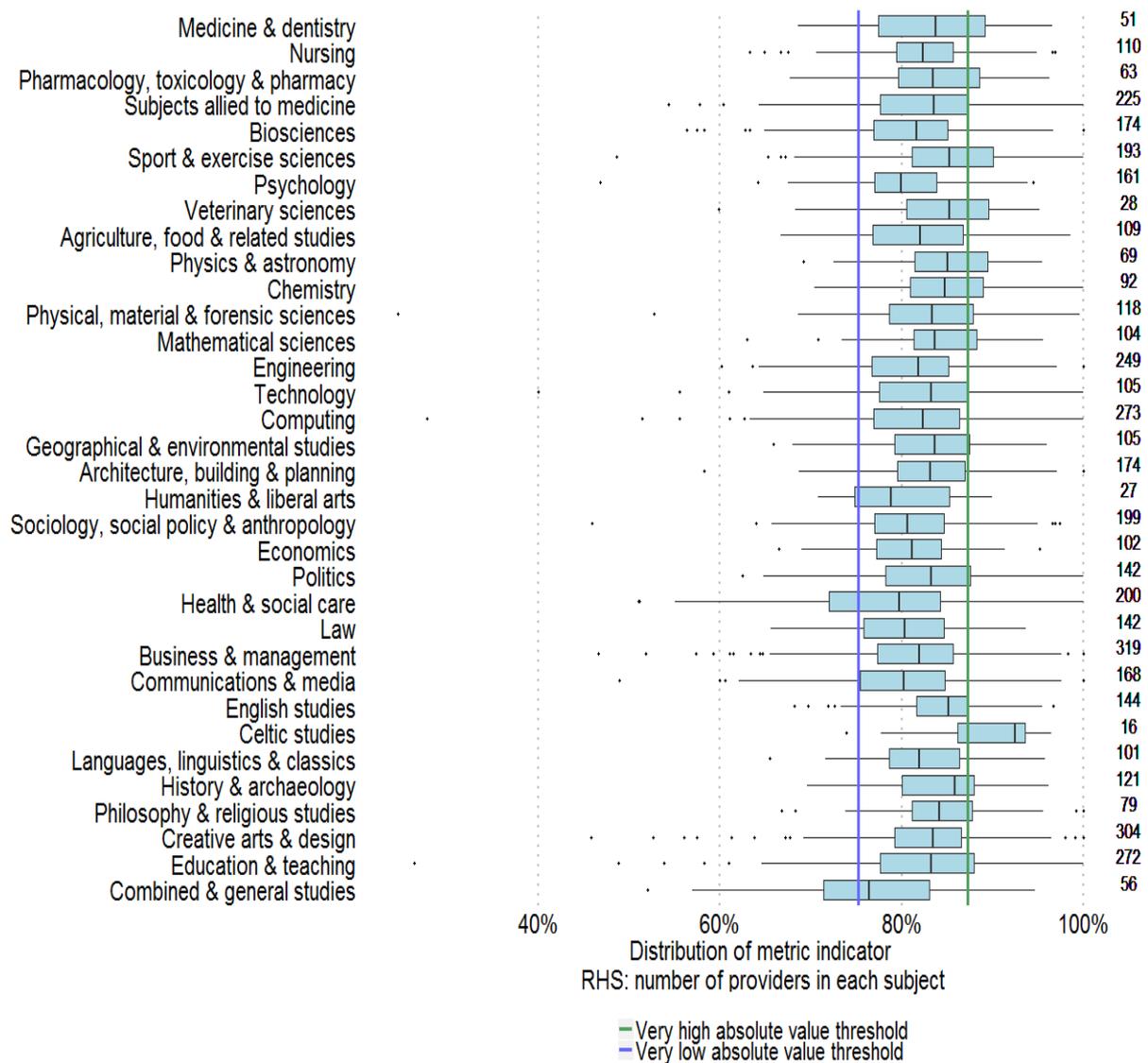
**Figure 11 Distribution of non-continuation indicators (full time)**



**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low absolute thresholds are from TEF data prepared for assessments in academic year 2017/18.

The Academic support metric is based on student’s responses to specific NSS questions which cover the NSS scale “Academic Support”. Figure 12 shows the distribution of this metric across providers for full-time students by the 35 CAH2 subjects<sup>36</sup>. The distribution of academic support indicators is similar for most subjects with indicators falling in-between the high and low absolute value thresholds for most providers.

**Figure 12 Distribution of academic support indicators (full time)**

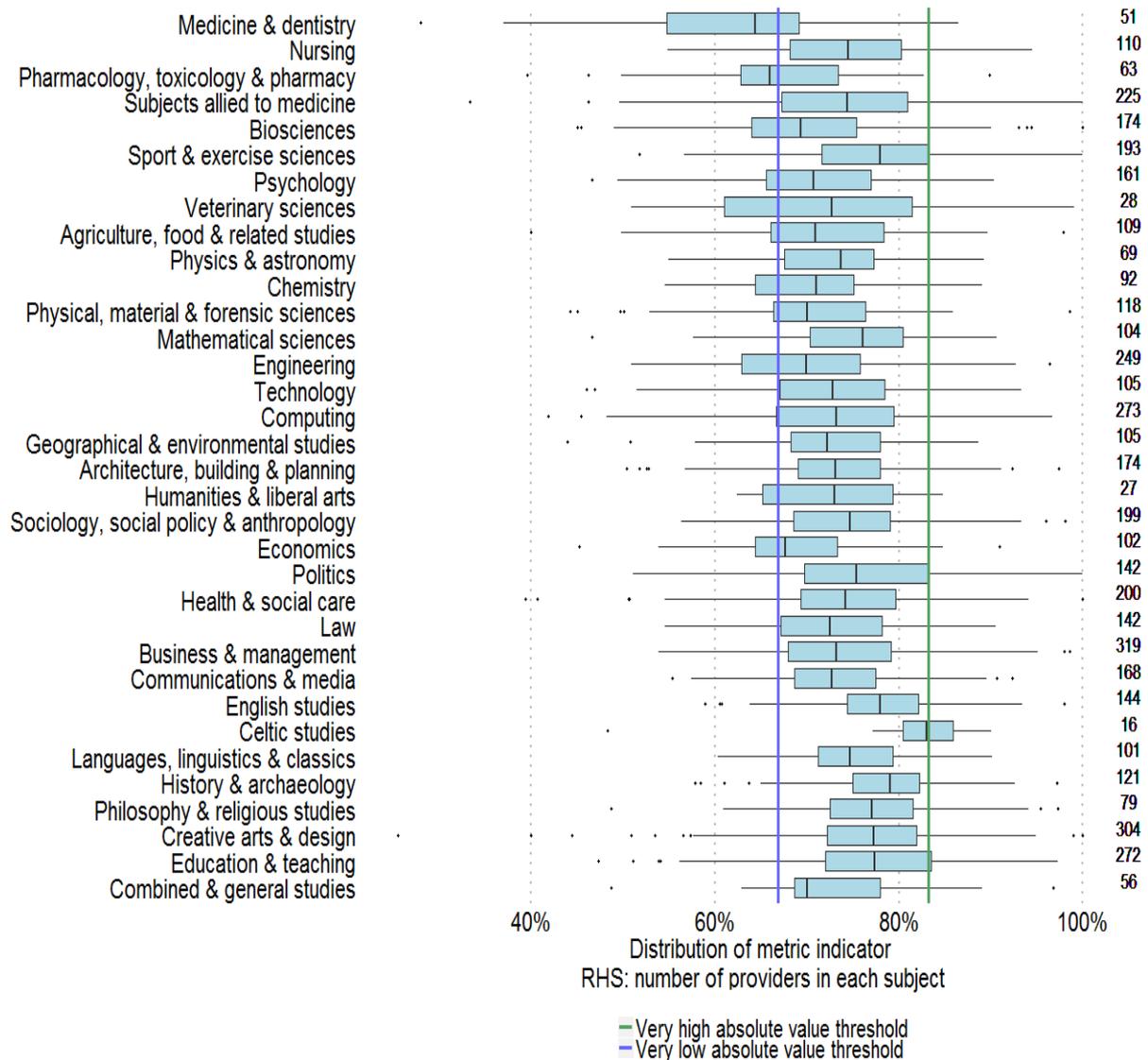


**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low absolute thresholds are from TEF data prepared for assessments in academic year 2017/18.

<sup>36</sup> See note to Figure 12 for coverage of providers.

Figure 13 shows the distribution of NSS 'Assessment and Feedback' indicators across providers for full-time students by the 35 CAH2 subjects was similar for most subjects<sup>37</sup>. There are two noticeable exceptions to this, however. Firstly, there is a smaller range for Celtic studies. Secondly, distribution for Medicine and dentistry sits lower along the indicator spectrum than most other subjects.

**Figure 13 Distribution of assessment and feedback indicators (full time)**

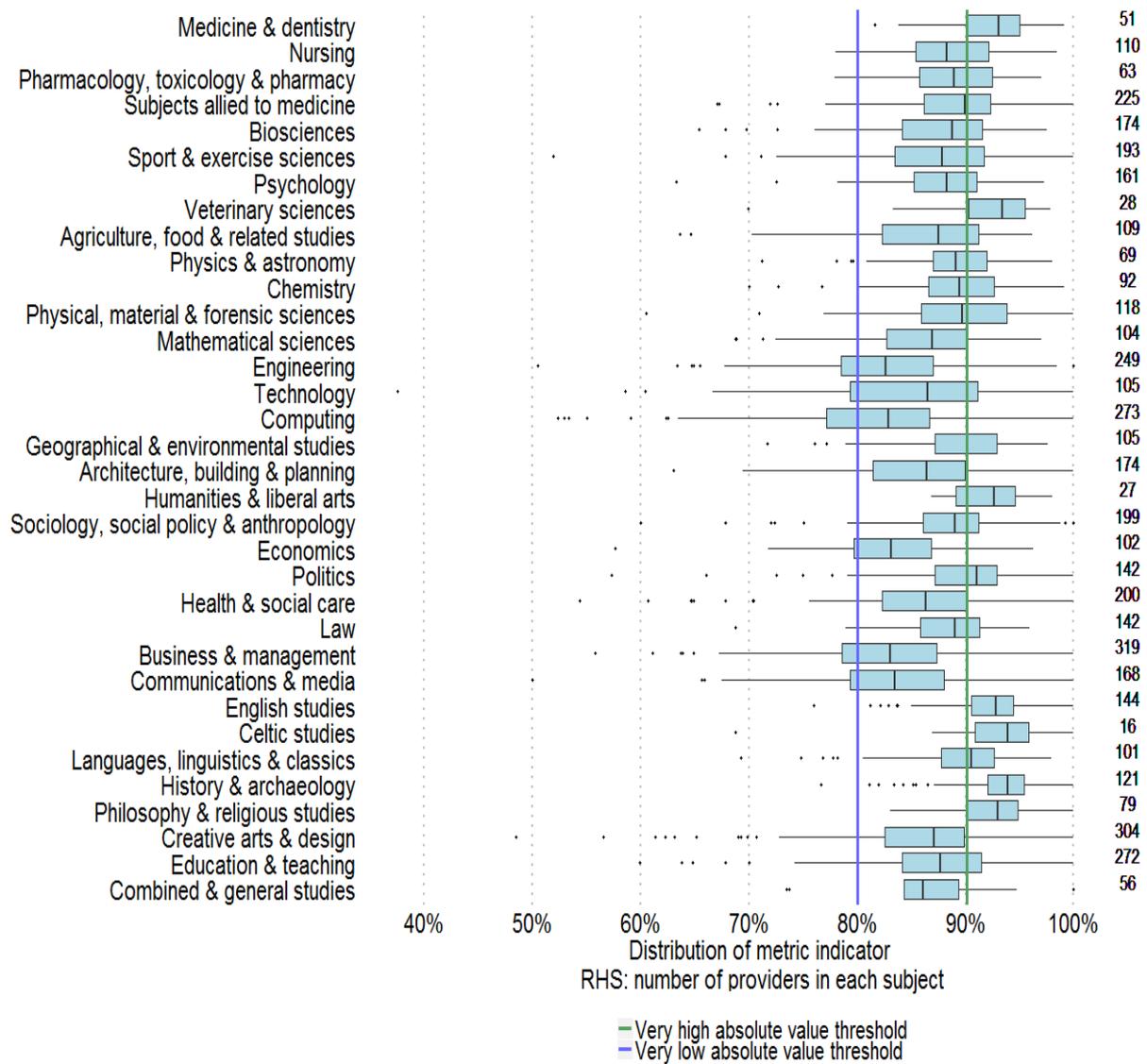


**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low absolute thresholds are from TEF data prepared for assessments in academic year 2017/18.

<sup>37</sup> See note to Figure 13 for coverage of providers.

Figure 14 shows the distributions of the NSS Teaching on my course indicators vary across the 35 CAH2 subjects although nearly all subjects had indicators over 80% for most providers<sup>38</sup>.

**Figure 14 Distribution of teaching on my course indicators (full time)**



**Notes:** Metric indicator data is from TEF data for assessments undertaken in academic year 2016/17 which includes providers that were in scope for TEF during the TEF assessments that took place in that academic year. This includes data from providers who were eligible for full and provisional TEF awards, and differs from the number of providers who received a TEF award in that assessment year. The very high and very low absolute thresholds are from TEF data prepared for assessments in academic year 2017/18.

<sup>38</sup> See note to Figure 14 for coverage of providers.

## Annex D: Student research outline

### Student TEF choices: The usability of subject-level classifications and teaching quality factors, to inform development of the Teaching Excellence and Student Outcomes Framework

The Government wishes to understand how subject-level TEF can best support student decision making. In particular, it wishes to explore two aspects of design:

#### Part 1 – Testing subject-level classifications

Given that there are presently 30,000 to 40,000 undergraduate courses available in the UK, TEF assessment at individual course level is not practical. Courses will therefore be grouped into subject-level classifications to make the data manageable. However, it will be necessary to ensure sufficient granularity to support informed student choices.

Currently the subject-level classification being considered is Level 2 of the Common Aggregation Hierarchy (CAH2), which classifies courses into 35 subject groups. Alternative classifications include a) the Level 1 Common Aggregation Hierarchy (CAH1) consisting of 23 groups, and b) 7 broad subject groups. Testing the usability of the CAH2 classification system against these two alternatives forms the first part of this research.

#### Objective

To identify which of three potential TEF subject-level classifications can be recommended as the most useable by students, by testing the classifications with undergraduate university applicants.

1a) Can participants use the classifications to quickly and correctly identify which subject category their preferred course falls under?

1b) Which subject classification is preferred?

1c) For the Common Aggregation Hierarchy 35-level classification, are there any subject classifications that could be improved, and if so, how?

1d) Are students aware of and using provider-level TEF, and is subject-level TEF preferable to provider-level TEF?

#### Sample 1

An estimated minimum sample would be 1,800 students. This shall include a representative sample of pre-applicants and applicants aiming to commence undergraduate courses in 2018/19. Purposive sampling shall also be undertaken to ensure there is sufficient representation of subjects to adequately test the three subject classifications.

Recruitment will be via the UCAS database.

## **Part 2: Understanding the Importance of Teaching Quality**

The TEF has been developed with significant input from the sector, including student representatives. Additionally, research has been undertaken which supports this development. The teaching quality factors for both provider and subject level TEF now need to be tested more systematically with students themselves to establish:

- a. the importance of each factor in selecting a higher education provider for undergraduate study, and
- b. the importance of each factor as an indicator of the quality of their undergraduate student experience.

The factors to be tested will include a broad range of factors that students might consider important when choosing a university. However, the research will only consider factors that are related to teaching quality in its broadest sense (for example, employability after graduation, contact time, calibre of teaching staff, facilities). It will not consider factors that are not associated with teaching (for example, the availability of bursaries or the quality of student accommodation). It is expected that there will be 20 factors selected for inclusion in this research.

Students will also be invited to add additional teaching quality factors that they may identify themselves, if they feel those factors are not already represented.

### **Objective**

Objective two: To understand students' views on which provider- and subject-level factors relating to 'teaching quality' in its broadest sense are important to students in terms of:

- 2a) selecting an HE provider
- 2b) the quality of their undergraduate student experience.

### **Sample 2**

A representative sample shall include around 1,000 applicants for UK HE in 2017/18, applying for courses starting in 2018/19, and 1,000 UK first and second year undergraduates who will have taken up their course in 2017/18 or 2016/17. Purposive sampling shall also be undertaken to ensure that all broad subject groups are adequately represented.

Recruitment for applicants will be via the UCAS database and for current undergraduates will be via Youthsight's Higher Education Student panel.

The sample shall achieve good representation of applicant and student characteristics in order to support the drawing of statistically robust results for different types of student and for demographic subgroups. These are likely to include course subject; international/home student; UK home region/Devolved Administration; HE provider of choice/attending; UCAS tariff (or expected grades); age; gender; mode of study; special educational needs.

## Annex E: GTQ example

The GTQ is calculated by multiplying taught hours by the appropriate weighting and summing the total across all bands. Table 9 sets out the staff/student ratio bands and the weightings for each band. These have been chosen to broadly reflect different class sizes (small group sessions, lectures, seminars etc.). The table also demonstrates how the GTQ is calculated.

**Table 9 Teaching intensity: staff/student ratio bands and the weightings**

Type of Contact (Staff/student ratio = X)	'Typical' students per staff member (R)	Weighting (W = 1/R)	Taught Hours (H)	Total weighting (T=HxW)
Taught: $X \leq 2$	1.5	2/3	0	0
Taught: $2 < X \leq 8$	5	1/5	6	1.2
Taught: $8 < X \leq 20$	14	1/14	2	0.143
Taught: $20 < X \leq 40$	30	1/30	2	0.067
Taught: $40 < X$	75	1/75	15	0.2
<b>Gross Teaching Quotient (<math>=\sum T</math>)</b>				1.61



Department  
for Education

© Crown copyright 2018

This document/publication (not included logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

visit [www.nationalarchives.gov.uk/doc/open-government-licence/version/3](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3)

email [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

write to Information Policy Team, The National Archives, Kew, London, TW9 4DU

About this publication:

enquiries [www.education.gov.uk/contactus](http://www.education.gov.uk/contactus)

download [www.gov.uk/government/consultations](http://www.gov.uk/government/consultations)



Follow us on Twitter:  
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:  
[facebook.com/educationgovuk](https://facebook.com/educationgovuk)