Annex C: Evidence from providers

This annex provides an analysis of feedback from the 50 providers that took part in the first TEF subject pilot, 2017-18. Feedback was captured throughout the process at briefing events, via a formal post-submission survey, and through in-depth interviews after providers received their ratings.

Methodology

1. Feedback was gathered from participating providers throughout the pilot through a series of events, comprising:

   - pre-application briefing events in December 2017

   - mid-application briefing events in January 2018

   - post-application briefing events in March 2018 (which was preceded by an optional informal online survey that was used to develop the agenda).

2. The pre-, mid- and post-application briefing events captured issues emerging throughout the process, including some we were able to respond to during the pilot. Providers were very engaged throughout the pilot and their views developed as they gained more experience of each model.

3. Event feedback was then used to inform the design of a post-application online survey that posed a series of qualitative and quantitative questions covering the whole experience of the pilot. After providers received their ratings in May 2018, a series of individual phone calls with each provider was held during June 2018, to explore providers' reactions to their ratings and get their final feedback on the pilot.

Views throughout the process: providers in both models

4. Providers participating in both Model A and Model B were asked throughout the process which of the two models they preferred: at each of the three events, in the provider survey and then for a final time in the phone calls.

5. From a starting point in December of 10 votes in favour of Model B, no votes in favour of Model A and one borderline (out of 11 respondents), their views steadily moved towards rejecting both models and favouring a hybrid combining the best features of both. In January there were four votes for Model A and four for Model B but six for a hybrid solution. In a dedicated workshop for participants in both models held during the post-application event in March there were 11 votes out of 12 for a hybrid model combining the full assessment of Model B with the submission formats of Model A: five-page subject submissions and 15-page provider submission. One single-subject provider chose to abstain.

6. To test whether the hybrid model would be more widely supported, a question on it was included in the provider survey (see paragraphs 34 to 36).

Views throughout the process: challenge

7. To assess the level of challenge of participating in the pilot while the experience was still fresh for providers, an informal survey before the March post-application briefing event asked how challenging providers had found it to make their submissions, and how the challenge compared with their expectation when they signed up to the pilot. The survey was optional, and was answered by 31 of the 50 participants. All 17 respondents to the question answered that it was challenging to make a submission under Model B, while 16 of 19 respondents had found it challenging to make a submission under Model A (participants in both models responded once for A and once for B).

Table 1: Model A – how did your experience compare to your expectations in September 2017?

Overall, how challenging was it to make a TEF application under Model A?	Much more difficult	Somewhat more difficult	About what I had been expecting	Somewhat easier	Much easier	Row totals
Most challenging	2	0	1	0	0	3
Somewhat challenging	3	6	4	0	0	13
Neither too challenging nor too unchallenging	0	0	1	1	0	2
Somewhat unchallenging	0	0	0	1	0	1
Least challenging	0	0	0	0	0	0
Column totals	5	6	6	2	0	19

Table 2: Model B – how did your experience compare to your expectations in September 2017?

Overall, how challenging was it to make a TEF application under Model B?	Much more difficult	Somewhat more difficult	About what I had been expecting	Somewhat easier	Much easier	Row totals
Most challenging	2	2	2	0	0	6
Somewhat challenging	2	5	4	0	0	11
Neither too challenging nor too unchallenging	0	0	0	0	0	0
Somewhat unchallenging	0	0	0	0	0	0
Least challenging	0	0	0	0	0	0
Column totals	4	7	6	2	0	17

Provider survey summary

8. Following the feedback collected from providers during briefing events, an online survey was issued to all pilot providers during April 2018. Of the 50 pilot participants, 48 responded; the two that chose not to respond were a small higher education institution and a further education college, both participating in Model B only. The survey collected detailed quantitative and qualitative feedback about various aspects of the pilot. Key findings are reported below.

General approach

9. In this section, providers were asked to indicate whether the fact that this was a pilot rather than a real exercise had influenced how they approached the submission, on a scale ranging from 'No influence' to 'Major influence', and to provide more detailed commentary on their approach.

10. Responses were mixed, with 18 suggesting it had had a minor influence (2 on a five-point scale where 5 is high) and 17 suggesting a moderate influence (4 on a five-point scale).

11. There appears to be no correlation between the level of influence indicated by a provider and the model they took part in, and no clear correlation between provider type and level of influence, with the exception of 'No influence' providers, a high proportion of whom are alternative providers of higher education.

12. In general, providers tried to approach the pilot as far as possible as if it were a real exercise. Several providers explained why they treated the pilot differently from a real exercise, citing the following main reasons:

    a. The short time between receiving the metrics in November and uploading submissions in February was highlighted as a constraining factor which inhibited providers' ability to do all they would have liked to do, especially as this period incorporated part of a teaching term and the winter break.

    b. The pilot was low-risk with no outward-facing consequences, resulting in less editorial oversight and less approval through formal governance structures.

    c. The pilot had to be de-prioritised against other more pressing areas of work.

Using and understanding the data

13. Of the 48 providers that responded, 34 agreed or strongly agreed that the subject-level metrics provided new data that enhanced how they understood their provision, suggesting that the subject-level metrics are of value. Single-subject providers, or those close to being single-subject, were more likely to disagree or be neutral, presumably because their subject metrics were the same or very similar to the provider-level. Two of the providers that disagreed were large higher education institutions, commenting that they already knew what the metrics were telling them. Several providers commented that the subject-level benchmarking is useful as it enables sector comparisons. An additional positive effect of metrics at the subject-level was that academic staff across the institution engaged more fully with data.

14. As expected, closed and new courses were identified as challenges not well served by the metrics, as were joint honours and interdisciplinary courses, courses with small cohorts and courses with non-reportable metrics. Providers also identified subjects where provision was split across two or more departments as challenging to write about (creative arts and design were frequently cited in this regard). Further education colleges appear to experience all these challenges more acutely than higher education institutions: they tend to have smaller cohorts leading to less reliable or non-reportable data, and more mergers that cause difficulty in interpreting which programmes contribute to the data. Another issue identified was the historic nature of TEF metrics, which means that any improvements providers make to their provision will not be visible in the metrics for several years.

15. Of the 48 responding providers, 30 said they 'often (in more than half of subjects)' or 'always (in every subject)' drew on quantitative data that was not covered by the metrics in their submissions, with a third (16 of 48) 'always' doing so. Additional data sources might best be described as 'our own internal datasets that complemented or conflicted with TEF data'; internal data of various types was frequently used (e.g. students' end-of-module evaluations and other internal surveys, internal teaching awards, external examiners' reports and PSRB accreditations, teaching evaluations), and providers' own more detailed analyses of datasets such as NSS or LEO results.

16. Industry maps were consistently judged to be not useful; 32 of 47 of responding providers, or two thirds, said they did not refer to them in a single submission. Just four of 47 used industry maps in more than half of submissions, and even two of those four commented that they were not very useful.

Staff engagement

17. Responses reveal a high level of academic engagement with the subject-level metrics: 24 providers (50 per cent) responded that academic staff had 'always (in every subject)' engaged with the metrics, and a further seven said that academic staff had engaged with the metrics 'often (in more than half of subjects)'. Only two providers responded that academic staff had never engaged with the metrics and five said that academic staff had engaged 'rarely (in less than half of subjects)'.

18. Obstacles to engagement this time were mainly the tight timescale of the pilot and staff's lack of familiarity with the process and metrics. However, it seems to have been harder for smaller providers to involve academic staff; a couple of them expressed doubt that they would be able to involve staff substantially more in future exercises.

Student engagement

19. Of 48 respondents, 40 per cent (19 providers) reported that they involved students in writing the subject-level submissions, and 63 per cent (30) reported that they involved students in writing the provider-level submissions. There was no clear relationship between the model providers were piloting and whether they had involved students with submission writing, and no clear relationship between provider type and student involvement. 30 per cent (15) providers reported that they did not involve students in writing either provider or subject submissions. However, half of these had interpreted 'writing' literally and actually reported some form of sharing and enabling comment from students via, for example, representation on TEF working

groups. Some of these providers explicitly indicated that they might have involved students in the submission writing in a real exercise, many of them also noting elsewhere in the survey, that in a real exercise 'tighter controls' and editorial oversight would be in place for all contributors.

20. Barriers to involving students in contributing to the submission included: confidentiality of the pilot and its associated metrics, active student policy against the TEF, tight timelines and resource constraints, and difficulties of navigating subject groupings in Model B.

21. Approximately 60 per cent of providers shared subject-level metrics with student representatives, although 40 per cent (19 providers) did not. No providers shared subject-level metrics data with the wider student body. Reasons cited for this were a perceived lack of student interest in them, a lack of time, and a concern that without full training and support they might have been misinterpreted, or the pilot's confidentiality breached. Most providers commented that they would plan to share metrics with students in a real TEF exercise.

Interdisciplinarity

22. About half (23) of the pilot providers that responded to the survey, and nearly two thirds of the universities, had a course mapped to two or more CAH2 subjects, a course mapped to one of the two multi-subject categories ('Combined and general studies' and 'Humanities and liberal arts') or both. For subjects that had been mapped to two or more CAH2 subjects, the majority of providers found it 'difficult' or 'very difficult' to write submissions, and the majority of respondents disagreed or strongly disagreed that they were able to adequately reflect their interdisciplinary provision.

23. Responses regarding the multi-subject categories were broadly split evenly between providers that found writing submissions difficult or very difficult and those that were neutral or found it easy.

Pilot documentation

24. Providers generally found our guidance useful, with over 70 per cent identifying the metrics guidance and rebuild document as 'quite' or 'very' useful and over 50 per cent agreeing that the submission writing guidance was quite or very useful. The teaching intensity guidance fared slightly less well: just under 50 per cent of providers agreed that it was quite or very useful. Providers made written comments offering specific suggestions for improvement that the OfS will use to develop the guidance in the second year of pilots.

Model-specific questions

25. Providers were asked a series of questions about their experience of the model they participated in. Providers that took part in both models answered both Model A and B questions. The results for Model A and Model B are here reported side-by-side to enable comparisons. Overall the results suggest that providers have no clear preference for one model over another, and are not strongly in support of either model.

26. Providers were asked about the model features that were designed to reduce burden. In Model A providers were asked to what extent they agreed that the by-exception model had facilitated a lighter-touch, risk-based approach. In Model B providers were asked to what extent they

agreed that assessing all subjects, but with a shorter provider submission and subject group submissions, facilitated a fuller assessment process while still being manageable. Opinions were split for the burden-reducing elements of each model, with slightly more disagreement than agreement apparent in each model. It should be noted that the difficulty for smaller providers in completing the significant workload of the pilot – whether participating in Model A, Model B or both models – was repeatedly commented on in the survey.

Figure 1: To what extent do you agree that this model enabled you to accurately represent provision at your institution?
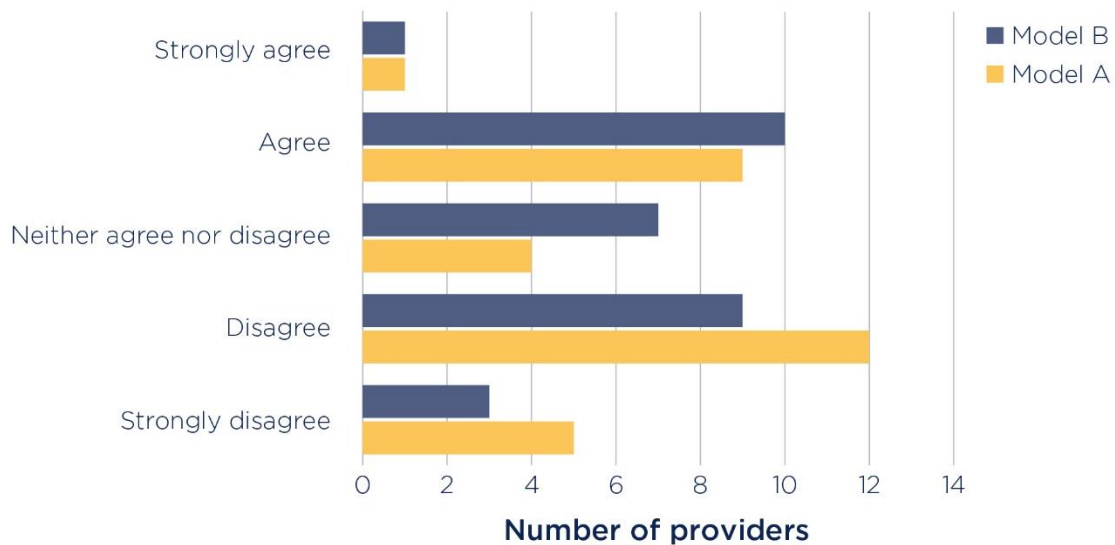
Figure 2: To what extent do you agree that the model facilitated a fuller assessment process while still being manageable?
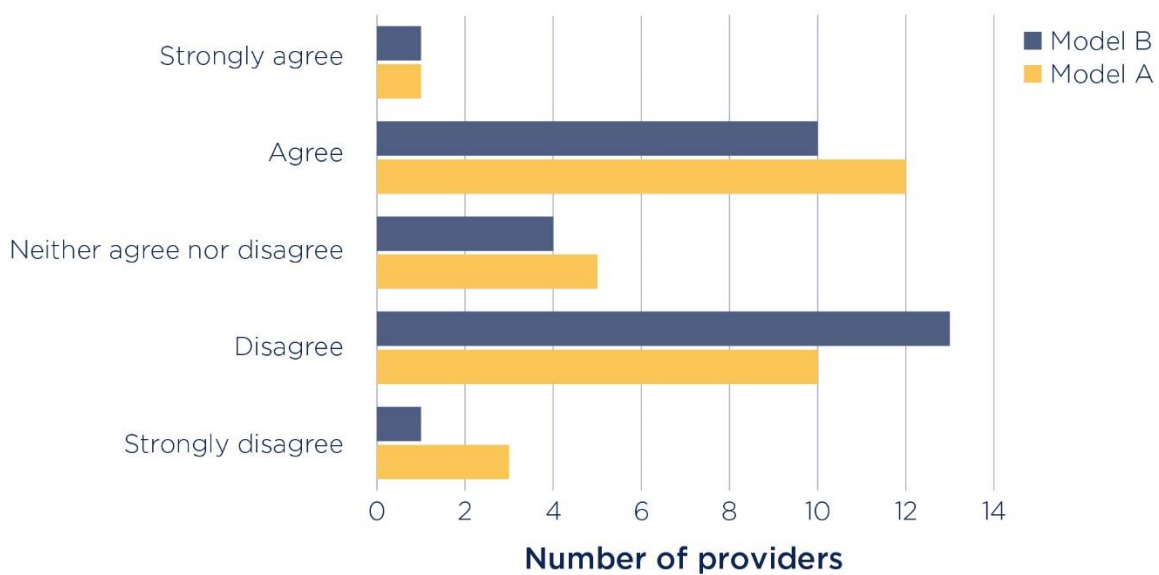
27. The key strengths of each model identified were the clarity and purpose of the subject-specific five-page submission Model A and the more comprehensive assessment approach in Model B. Five providers that participated in Model B (all multi-faculty higher education institutions) explicitly suggested that individual submissions for every subject at CAH2 level would be better than the current Model B subject group style submission.

28. In Model A, those disagreeing with the by-exception method tended to object in principle to this approach. Several expressed concerns about the fact that a provider rating might change after the exceptions were generated, questioning whether this approach would be fair or robust in a real exercise, and noting that the workings of the exception algorithm could also be difficult to understand, and difficult to explain to submission authors. Opinion was more split on whether the right exceptions had been identified in Model A, with 13 providers agreeing or strongly agreeing, 10 disagreeing or strongly disagreeing, and seven neither agreeing nor disagreeing.

29. In Model B, it was felt that subject group submissions would provide poor information for students, because of their aggregated nature and that the tight page limits that did not give enough space to do justice to any subject. Most respondents (21 of 29) also disagreed with restricting the provider submission to three criteria as this created an artificially narrow focus. The ability to move subjects between groups in Model B, however, was appreciated (and used by 55 per cent of respondents). All of those providers that moved a subject between groups suggested that the change better reflected the internal organisation of the provider's faculties and departments.

Subject classification

30. The survey asked about the use of CAH2 to categorise subjects. Although the quantitative survey responses suggest a wide spread of opinion about CAH2 subject classifications, there is an underlying consensus that an amended CAH2 represents the best available option for TEF. There is little support for any alternative classification system.

Figure 3: To what extent do you agree that categorising subjects on the basis of the CAH2 allowed you to represent your provision well?
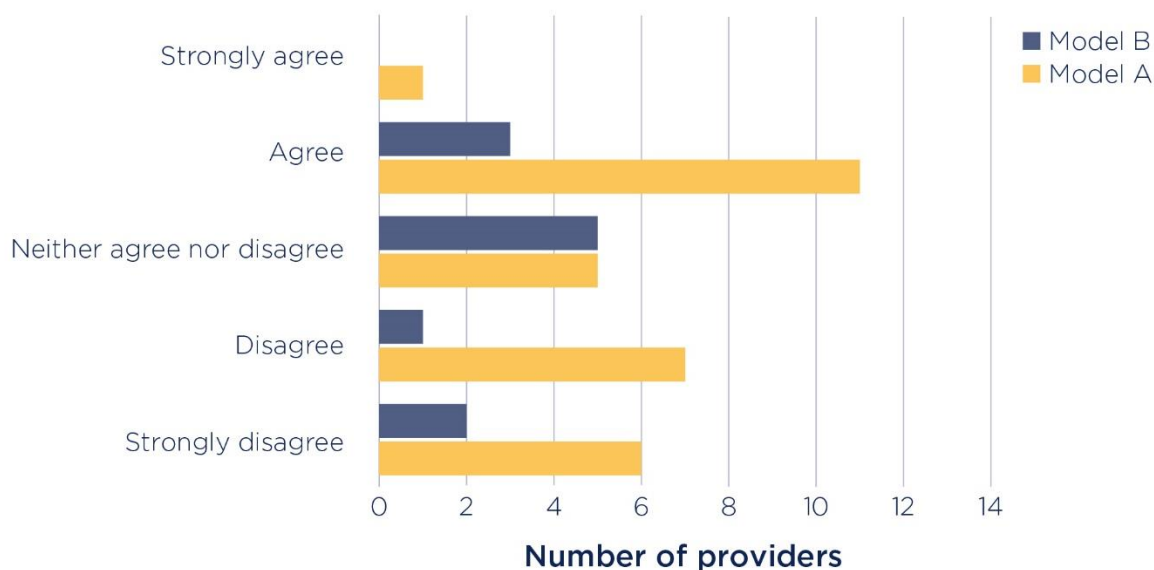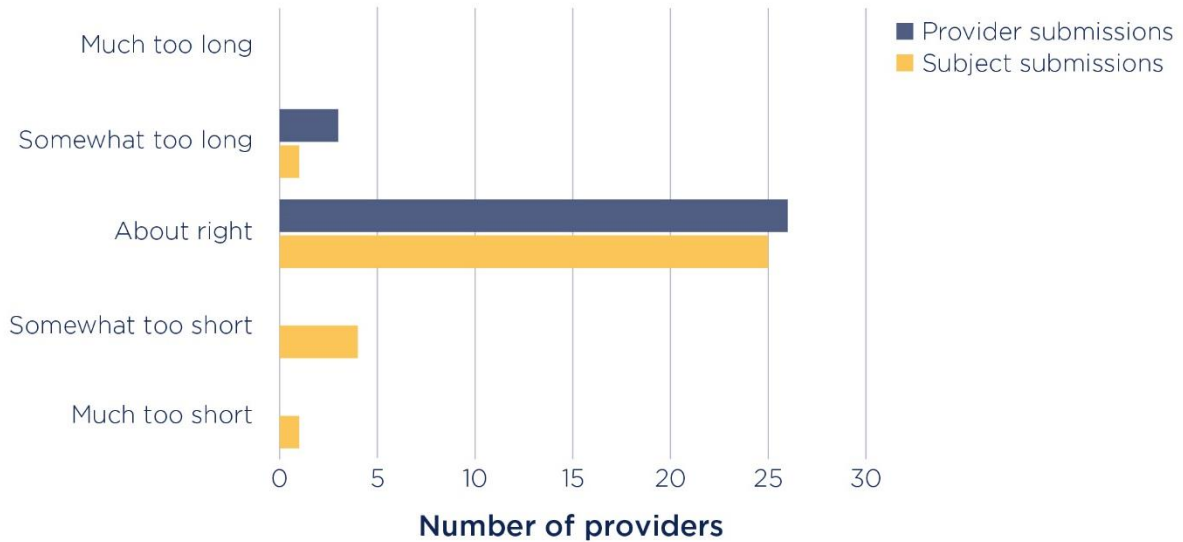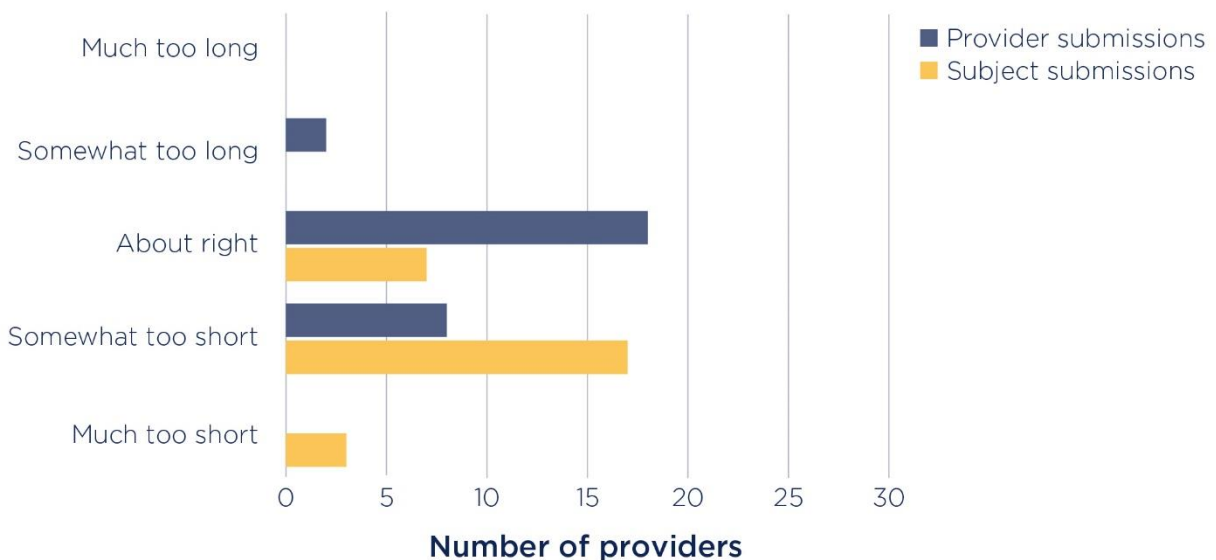
Page length

31. The survey also asked whether the page length of provider and subject-level submissions were right. As illustrated in the figures below, the 15-page provider submission and five-page subject submissions in Model A seem to be about right.

Figure 4: Model A – appropriateness of page lengths

Figure 5: Model B – appropriateness of page lengths

Non-reportable metrics

32. Non-reportable metrics were generally rated as difficult to write about, although there is almost no difference between the models: in Model A 13 found writing about non-reportable metrics difficult, four found it easy and five were neutral, while in Model B 13 found it difficult, three

found it easy and six were neutral (the other respondents either skipped the question or responded that they had no non-reportable metrics).
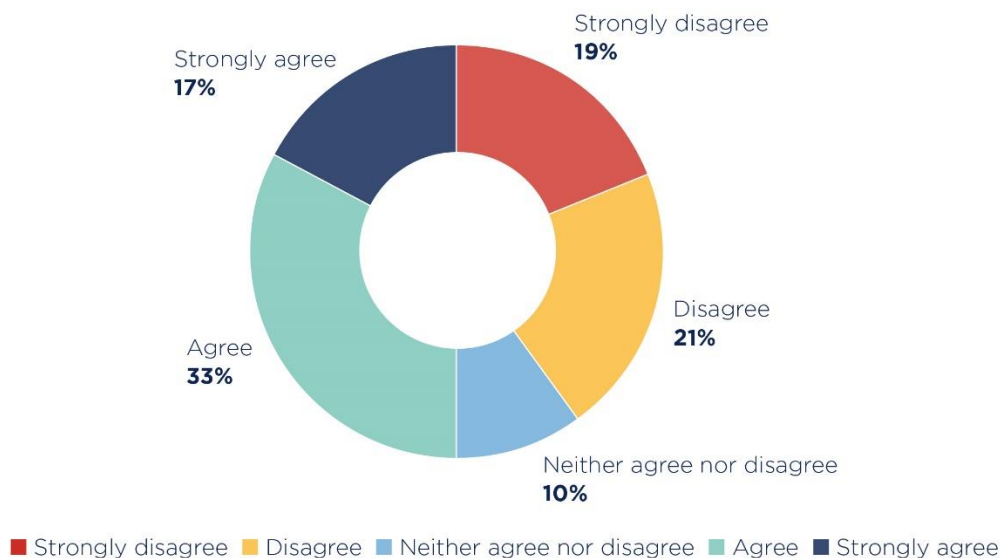
Providers in both models: comparative view

33. As described above, views from providers that took part in both models point towards rejecting either model in its current form and exploring a more comprehensive approach, drawing on the best elements from both models. However, when they were asked to choose between Model A or Model B only in the provider survey, six providers preferred Model B, four preferred Model A and two had no preference. (One provider later changed preference from Model B to Model A in the phone calls: see paragraph 60).
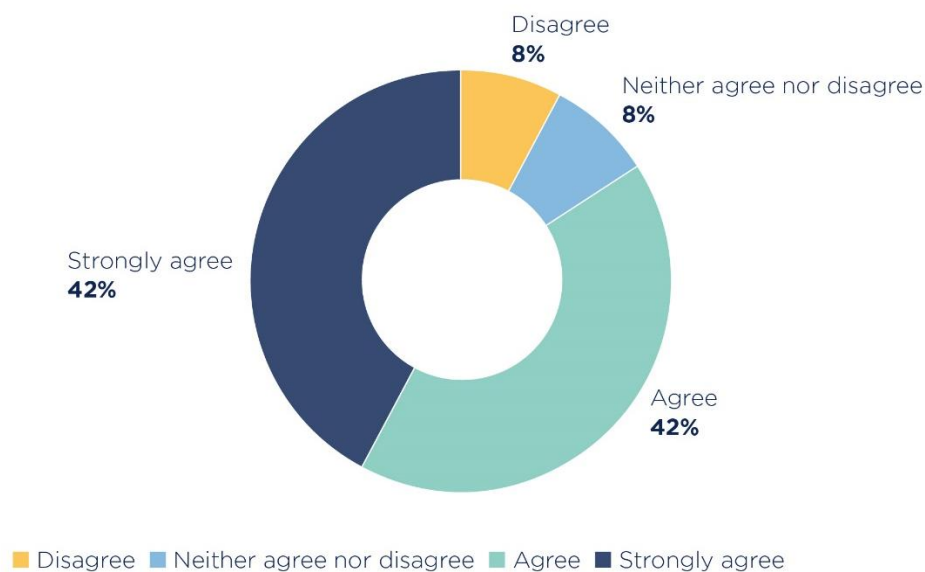
Model refinements

34. There is reasonable support for the idea of a more comprehensive assessment approach comprising a provider submission responding to the full set of criteria and a subject submission for each subject with a metrics workbook. The survey showed that among all pilot participants, 50 per cent were in agreement, though a sizeable minority (40 per cent) disagreed with the suggestion, in the main smaller providers concerned that a more comprehensive assessment would increase providers' workload. A few other respondents disagreed as they opposed the idea that submissions would have to cover all 10 criteria rather than being able to respond to those they wish, as is currently the case.

Figure 6: Level of support for a more comprehensive model (all providers)

Strongly disagree
19%

Strongly agree
17%

Disagree
21%

Neither agree nor disagree
10%

Agree
33%

■ Strongly disagree ■ Disagree ■ Neither agree nor disagree ■ Agree ■ Strongly agree

35. However, support for the more comprehensive approach among the 12 providers that participated in both models was unambiguous: 10 agreed or strongly agreed with the proposal, one disagreed and one was neutral. It should also be noted that of the four further education colleges and alternative providers that participated in both models, for whom burden concerns might be expected to be highest, three preferred Model B in the provider survey.

Figure 7: Level of support for a more comprehensive model (providers participating in both models only)

Disagree
8%

Neither agree nor disagree
8%

Strongly agree
42%

Agree
42%

■ Disagree ■ Neither agree nor disagree ■ Agree ■ Strongly agree

36. Other suggestions for refining the existing models included not generating Model A exceptions until after the provider-level rating is finalised, and more systematically making information around provider context available to subject panels.

Teaching intensity

37. Collecting data for the provider declaration was difficult or very difficult for most pilot participants and fewer than half of them referred to teaching intensity data in their submissions. 24 disagree or strongly disagree that Gross Teaching Quotient (GTQ) or a similar measure should form part of any future exercise (compared with 15 who agree or strongly agree), and 27 disagree or strongly disagree that students' own reported contact time should form part of any future exercise (compared with 12 agreeing or strongly agreeing). That said, there is nuance in the results: providers show more agreement with the external visits and online teaching measures than with the GTQ and student survey, and the majority of further education colleges are in favour of provider-declared teaching intensity measures being included in future exercises.

38. For a fuller description of the survey results for teaching intensity, see Annex E.

Post-results phone calls summary

Methodology

39. Providers received their own ratings confidentially at the end of May 2018. Providers did not receive any information about the ratings of other participating providers.

40. Following the release of ratings, members of the TEF team at the OfS conducted half-hour phone calls with each provider to get their final views on the pilot after they had received their ratings. The majority of the phone calls were carried out over a three-week period in June.

41. The phone calls were structured around an agenda that had been shared with providers ahead of time, and explored three broad questions:

    a. How providers felt about the ratings they had received.

    b. What impacts they believed these ratings might have for applicants, students and the provider if they were not bound by the confidentiality agreements of the pilot.

    c. If there was any other feedback providers wished to give or any earlier feedback they wished to change their minds about in the light of their ratings.

42. Providers in both models had been asked at intervals throughout the process which of the two models they preferred, and were asked again during these phone calls what their preference was. Findings are summarised at paragraphs 59 to 61 below.

Reactions to ratings

Reactions to provider ratings

43. Most of the provider-level ratings were as providers had been expecting and these providers were not surprised by their results. Providers that were surprised or disappointed by their ratings tended to raise questions about the balance of metrics to submissions in the assessment process and the role of the subject-based initial hypothesis in Model B (including for one provider rated Silver in A but Bronze in B). There were also several comments that the more limited ten-page Model B provider submission had made it difficult to know what to include and exclude in the submission.

Reactions to subject ratings

44. Subject ratings were in most cases agreed to be good fits, and again there were not many surprises. Providers in general already knew where their strengths were and which areas they needed to work on.

45. We collected feedback on how providers felt about some subjects receiving 'No rating', where applicable, as providers were unaware of the 'No rating' option prior to making submissions. The response was on the whole very positive: in most cases providers could easily understand why a subject had not been rated and agreed with the decision. Only five of the 16 providers (three higher education institutions and two further education colleges) that mentioned no-ratings expressed disappointment or frustration.

46. In some cases providers were frustrated by the lack of reportable metrics for non-rated subjects because of, e.g., small cohorts, part-time provision with poor DLHE response rates, and short courses that were not eligible for NSS. These issues disproportionately affect further education colleges. One further education college agreed with the principle of 'No rating' for non-reportable subjects, but expressed concern about how non-rated subjects would influence the provider rating. Four further education colleges felt that providers with some non-reportable subject metrics should be able to make a case using internal data and receive a rating.

Model A non-exception subjects

47. Some providers (including two further education colleges) were happy for non-exception subjects to inherit the provider rating, citing reduced burden, but several wished they could

have put forward some additional subjects for assessment. The provider that moved up from Bronze to Silver at provider-level was very happy with non-exception subjects inheriting the Silver rating. However, a couple of Gold-rated providers suggested that although they were glad to receive inherited provider ratings for non-exception subjects, the provider rating might not be representative of provision in all subjects. One provider pointed out that one of its subjects with no reportable metrics was able to inherit the provider-level Silver, which seemed inconsistent when another of its non-reportable subjects that was assessed had received no rating.

48. It was also noted that some subjects that were selected for assessment, even non-exception subjects chosen as control cases, had received a rating different from the provider rating (whether higher or lower). Model A was therefore widely recognised as flawed, even by providers that benefited from this design feature.

Perceptions of the potential impacts of ratings

Potential impacts for applicants

49. At this very early point, providers' sense of the impacts of TEF awards is based on perceptions, and seven providers observed that it is too soon to tell what effects TEF will have on applicants. Of 50 providers, 13 took the view that applicants may not be much influenced by TEF results and will choose their higher education provider for other reasons (e.g. locality, reputation).

50. A couple of providers raised concerns that some applicants might be put off of applying to Gold-rated institutions or subjects, and that this might have negative impacts on the widening participation agenda. One of these providers had arrived at this response after undertaking some research with its own current students; the other may have been influenced by other published research on the impacts of TEF awards on applicants[1].

51. A more common response, however, was the view that applicants may be attracted to higher-rated subjects and may seek to avoid Bronze. Several providers said they would not advertise Bronze awards. Several providers raised a specific concern about the effect of a TEF Bronze on international applicants. A future lack of international students could have significant impacts for some providers.

52. Having subject ratings in addition to a single provider rating creates some new challenges in explaining and managing ratings that are different from the provider-level, particularly when the subjects are rated lower. Providers were concerned about how both applicants and current students might interpret an outcome of 'No rating'.

Potential impacts for students

53. The consensus opinion was that current students were likely to be dissatisfied with a Bronze rating and reassured or pleased by a Gold. Two providers raised the possibility of a 'halo effect' arising whereby students in Gold-rated subjects might feel more positive about their courses

---

[1] See, for instance, 'Teaching excellence: The student perspective – Research commissioned by a consortium of students' unions', trendence UK, November 2017, https://studentsunionresearch.com/.

and students in Bronze less positive, and respond to the NSS accordingly, thereby creating a TEF-NSS feedback loop.

54. A more widely-shared concern was what TEF ratings might do to current students' employment prospects, with several providers fearing that employers might stop recruiting from Bronze providers and subjects. Related to this is a concern about ratings that shift from year to year: how would students feel who had entered a Silver-rated course but graduated from a Bronze-rated one, and what effect would this shift have on student employment prospects? This again has the potential to loop back into TEF metrics.

Potential impacts for the provider

55. There was overwhelming agreement that the ratings and subject-level metrics would be used to drive internal enhancement and were already having positive effects (some providers identified that they were particularly useful for weaker departments). In several cases TEF metrics and, in one case, even TEF-style submissions were being added to internal quality processes. Where providers disagreed that the TEF would drive internal enhancement, the most common reason cited was that aggregation to subject-level does not provide information granular enough to be useful.

56. Several providers anticipated that subject ratings would increase competition within and between departments, although this was not necessarily seen as a bad thing. A couple noted that subject ratings would enable management decisions that might lead to closures or staff cuts.

57. It was noted that some staff were likely to become demotivated if they received Bronze ratings. Some providers suggested this had already happened following the TEF Year Two results. Other providers took the view that Bronze ratings would act as an encouraging rather than demotivating impetus for change (see paragraph 55).

58. Three of the eight further education colleges (including one that had been rated Bronze) commented that TEF ratings had raised the profile of higher education in further education colleges, and they were very positive about the value of TEF participation for them.

Providers in both models

59. As noted at paragraph 35, providers in both models rejected both Model A and Model B and expressed a strong preference for a hybrid model that combined both. However, they had been asked throughout the process which of the existing models they preferred, so their views could be charted over the course of the pilot, and they were asked this for a final time in the phone calls.

60. Only one provider reversed its opinion having received its ratings, from a preference for Model B to a preference for Model A, on the grounds that Model A had better reflected its provision.

61. At the end of the pilot process, when asked which of the *existing* models they preferred, this eft five providers preferring Model B, five preferring Model A, and two with no preference either way. However, even when asked which of the two existing models they preferred, five providers observed again that they would really prefer a hybrid model combining the holistic coverage of Model B with the provider and subject submission styles of Model A.

Other comments

62. Very little new feedback was offered at this stage; providers agreed that they had already represented their views accurately and fully in the post-pilot survey. The main point that emerged was that providers would have liked to receive individual feedback on their subject ratings (e.g. statements of findings), which the OfS intends to pilot next year. Providers in the devolved administrations raised the question of whether TEF participation would represent value or value for money for them and anticipated that not all their compatriots would conclude that it did.

63. In general providers were positive about the pilot, agreeing that it had been a constructive experience. Many expressed an interest in participating again next year.