Annex E: Teaching intensity

This annex gives an overview of the development and application of an exploratory teaching intensity metric in the pilot. It discusses feedback on teaching intensity from providers and panel members, and includes an analysis of the data that was captured.

Background

1. As was set out in the 'Teaching Excellence Framework Year Three subject-level pilot specification' published on 20 July 2017[1], the first subject pilot incorporated an exploratory pilot of a teaching intensity metric. Data on contact hours, class sizes and type of provision (face to face, online, or external visits and work-based learning) was collected from participating providers. This data was used together with student satisfaction information, collected directly from students at the participating providers, to develop teaching intensity metrics. The teaching intensity information was then made available to providers to comment on in their subject pilot submissions and also to panel members for use in assessment.

2. Fuller details of the teaching intensity pilot, including the guidance made available to pilot providers, were published on HEFCE's website in January 2018 to inform the wider higher education sector's responses to the DfE's consultation on subject-level TEF. This guidance is now available on the OfS website[2].

3. As the teaching intensity element was 'a scoping pilot with the intent of exploring the feasibility and usefulness of collecting and assessing this data'[3], the data collection for the first year of pilots was designed with the aim of using the simpler cases to test the broad principle and feasibility of the general approach. For instance, the provider declaration data collection was limited to students on full-time single-subject courses in their first, second or third taught programme year. As data collection was kept relatively simple during the first year of piloting, it did not accurately reflect some kinds of provision, for instance four-year and interdisciplinary courses. A number of other exclusions were made, which are detailed in full in the teaching intensity guidance for providers[4].

4. Teaching intensity was piloted in only five subjects: business and management; creative arts and design; engineering; history and archaeology; and nursing. These were not the five

[1] See https://www.gov.uk/government/publications/teaching-excellence-framework-subject-level-pilot-specification.

[2] 'Measuring teaching intensity: Guidance for providers', Office for Students, 30 January 2018, www.officeforstudents.org.uk/advice-and-guidance/teaching/piloting-tef-at-a-subject-level/further-technical-information.

[3] 'Teaching Excellence Framework: Year Three subject-level pilot specification', DfE, 20 July 2017, https://www.gov.uk/government/publications/teaching-excellence-framework-subject-level-pilot-specification, paragraph 209.

[4] 'Measuring teaching intensity: Guidance for providers', paragraph 11.

subjects originally identified in the specification[5]; changes were made to better reflect the provision offered by pilot participants while ensuring the subjects chosen represented a broad range of pedagogical approaches.

5. The timeline for delivering teaching intensity is given in Table 1.

Table 1: Teaching intensity timeline

Activity	Date
OfS design phase	July to November 2017
Data collected from providers	November 2017 to 12 January 2018
Student survey collecting data from providers' students	November 2017 to 12 January 2018
Data verified by the OfS and used to generate teaching intensity metrics	12 January to 9 February 2018
Teaching intensity metrics released to providers	9 February 2018
Providers uploaded pilot submissions (incorporating teaching intensity metrics if they chose to)	26 February 2018
TEF subject pilot assessments	March to May 2018
OfS analysis and evaluation phase	March to July 2018

Pilot data collections

6. Data for the teaching intensity metrics was collected in two parts: the provider declaration of teaching intensity and the student survey. The measures piloted were based on those outlined in the specification[6], but developed and refined by the OfS (see paragraph 14). Of the 50 pilot providers, 46 were in scope to complete the provider declaration of teaching intensity and 49 had courses in scope for the student survey. All providers returned teaching intensity data as required.

Provider declaration

7. Providers returned data in three categories:

- Gross Teaching Quotient (GTQ)
- external visits and work-based learning (WBL)
- online teaching.

[5] The five subjects originally identified were creative arts and design, history and archaeology, law, nursing, and physics and astronomy.

[6] 'Teaching Excellence Framework: Year Three subject-level pilot specification', paragraphs 216, 261-264, 267-268 and 270-273.

8. **Gross Teaching Quotient** recorded scheduled teaching activity provided directly by members of staff in real time, either face-to-face or online, reported in hours per year and weighted by the student-staff ratio of each taught hour.

9. **External visits and work-based learning** recorded scheduled learning activity in taught years of study that occurred outside usual face-to-face teaching. This typically involved students in activity supervised by staff or appointed representatives, for instance on a placement, and was reported in days per year.

10. **Online teaching** recorded the time spent by staff in facilitating online learning when they were not necessarily online at the same time as the students (asynchronous online teaching), because this asynchronous teaching could not easily be recorded as part of GTQ. This was reported in hours per year. Online teaching excluded the time students spent in independent learning online and excluded staff's preparation time; both of these exclusions were to ensure consistency between activity included under online teaching and teaching included under GTQ.

11. Providers were asked to return as much activity as possible under GTQ, reserving the other two categories for activity that clearly did not fit it, with a view to keeping the data collection as simple as possible.

12. Providers recorded contact hours and class size information for each full-time undergraduate module in scope for each subject in Excel workbooks supplied by the OfS, and uploaded them to the TEF extranet. The provider declaration was ordinarily made based on HESA data returns from 2015-16 unless the provider had requested to return based on 2016-17 data.

13. Hours recorded reflected hours that students were expected or scheduled to attend rather than actual attendance.

14. As noted above, the measures piloted had to change slightly from those proposed in the specification for practical reasons; in particular, weighting the GTQ by student module-to-course ratio because OfS analysts identified that the original formula might have led to misrepresentation. Another change involved recording GTQ in weighted hours per year rather than weighted hours per week. The specification had identified an 'e-learning typology', so the online teaching metric was developed by the OfS before implementation. For fuller details of the measures piloted, and fuller details of the adaptations made to the pilot specification, please see Annex E of the full teaching intensity guidance[7].

Student survey

15. The teaching intensity student survey was sent to all students on courses in scope at each provider. The aim was for student survey data to be combined with provider declaration data to create a metric indicating not just how much one-to-one equivalent teaching was provided, but how satisfied students were with the teaching they received on their course. In the survey, students were asked to identify the average number of hours per week spent in contact with teaching staff and in independent learning, before responding to these four statements about their experience of teaching during the past term or semester (autumn and winter 2017):

[7] www.officeforstudents.org.uk/advice-and-guidance/teaching/piloting-tef-at-a-subject-level/further-technical-information/.

a. ’There is enough teaching (face to face and/or online) to support my learning.’

b. ’The balance between teaching and independent study is right for my course.’

c. ’The amount of teaching on my course meets my expectations.’

d. ’Overall I am satisfied with the amount of teaching offered on my course.’

16. These questions were cognitively tested with student focus groups before use.

17. The OfS created one survey per provider, which providers then forwarded to their students on the relevant courses. Courses and students in scope were defined more broadly for the student survey than for the provider declaration: all students on a course were in scope, including part-time students, joint honours students and those on higher apprenticeships or accelerated degrees. The survey was sent to students in their first, second and third taught programme years in 2017-18.

18. To enable benchmarking of the metric, providers separately returned to OfS lists of all students to whom the survey had been sent so that OfS analysts could match data from survey respondents with providers’ HESA or Individualised Learner Record (ILR) returns, as these data returns show students’ demographic characteristics. However, the survey suffered from poor response rates and was over-represented in certain demographics (see paragraph 44), which meant it was not appropriate to benchmark the metric.

Creating the metric

19. Following the data collections, the OfS generated a teaching intensity metric for each provider per subject. The metric reported for each of the first, second and third taught years of study the mean, lower and upper quartiles of the amount of taught activity for a student in the subject, based across all students on courses within the subject for each of the three provider-declared categories (in one-to-one equivalent hours per year for the GTQ, in days per year for external visits and work-based learning, and in hours per year for online teaching) and the percentage of students satisfied with the amount of teaching they received on their course. Students were asked to identify the subject of their course, which allowed their response to be allocated to a subject. It also reported the student survey response rate for the provider. Note that the provider declaration and student survey referred to different student cohorts (respectively 2015-16 and 2017-18).

20. Providers received both their own teaching intensity metrics and some comparison information developed from the responses from all pilot providers to enable them to incorporate teaching intensity into their subject submissions if they wished. The teaching intensity metrics were added to each provider’s metrics workbook for the relevant subject and made available to panel members for use in the pilot assessments. Panel members also had access to the same comparison information as providers.

Operational issues

Provider declaration

21. Following their submission of the provider declarations, OfS analysts worked with providers directly to understand how reliable the data was, and also sought some optional feedback on the assumptions from each of the subject-pilot participants to help us understand the results they reported. Providers' responses to teaching intensity were more fully explored later in a survey sent to all pilot participants following the submissions (see paragraphs 25 to 33).

22. There was substantial variation in approaches and assumptions employed by providers in data declarations, typically driven by internal data availability or the lack thereof (often meaning a reliance on the record-keeping from previous years and the goodwill of individual academics). Because of these factors, providers had to make quite varied levels of assumptions when doing the data returns, including varied levels of estimation between different subjects, reducing the reliability of the data collected. The analysis that the OfS has undertaken of these returns is heavily impacted by these data quality concerns.

Student survey

23. The student survey received 4,880 complete responses (in which all questions were answered fully) of a potential 113,000 responses, which is a 4.3 per cent response rate. There were additionally 770 partial responses, in which only some questions were answered. These partial responses have not been included in the calculation of the metric. The partial responses have been analysed for patterns and biases causing students not to complete the survey, but none have been observed (see paragraph 45). As with the provider declaration, analysis of the student survey returns is heavily impacted by these data quality issues.

24. Student survey fatigue and the survey's timing over the winter holidays have been suggested as possible drivers of the low response rate. The latter issue was recognised in advance, but the tight timeline of the pilot made it impossible for the survey window to be extended. There was also a more specific issue of providers reluctant to send the survey to their final-year students out of concern that engaging with the teaching intensity student survey could have a negative impact on those students' subsequent NSS responses. One further education college and seven higher education institutions chose not to send the survey to their final-year students.

Provider feedback

25. Of the 46 providers that completed the provider declaration, 45 gave feedback on the process in the post-pilot survey. Of these 45:

    - 37 (82 per cent) found it difficult or very difficult to collect the data required for the GTQ
    - 31 (69 per cent) found it difficult or very difficult to collect the data required for external visits and work based learning
    - 27 (60 per cent) found it difficult or very difficult to collect the data required for online teaching

- 25 (56 per cent) found it difficult or very difficult to administer the student survey.

26. A key driver of this difficulty seems to be that many providers do not keep internal records of this data in the format required for the declaration, and therefore had to create it manually for the pilot, which was very-time consuming. One provider commented:

> 'Like many of the providers that took part in the teaching intensity pilot, a lot of the data had to be collated manually using a combination of business object [business intelligence] reports, "local" departmental knowledge and a degree of estimation. This approach would be unsustainable if the data we had to collect for the teaching intensity pilot was required for the university as a whole.'

27. It is likely that providers would have to invest in new data collection and management information systems in order to return teaching intensity data if it were to be more widely required in future years, and the initial costs of this would be high. The costs of returning teaching intensity data would also be substantially increased if the collection were extended in future to include more complex cases, such as part-time provision, that were excluded from this year's pilot. Considering the difficulties encountered by a small number of providers making a limited data return in this pilot, there would need to be very significant development and expert input before the OfS could create a reliable way of collecting teaching intensity data that is inclusive of all types of provision.

28. Data returned by providers in the cost survey confirmed the impression that collecting and returning the teaching intensity data had been a highly burdensome task. Individual providers' spending on the provider declaration of teaching intensity varied considerably, from just over £100 (for a provider with one teaching intensity subject) to almost £24,000 (for a provider with five), depending on the level of data available internally, the level of assumption made, and the amount of time spent checking the data. The mean spending on the provider declaration of teaching intensity in this year's pilot has been estimated as £3,479 per provider, and the mean spending per teaching intensity subject (using data returned by providers with five teaching intensity subjects) at £1,175. It is tentatively estimated that if a teaching intensity data collection similar to the one in this year's pilot had been carried out by all eligible higher education providers in every subject, it would have cost the higher education sector about half as much again as the total cost of the pilot exercise as a whole.

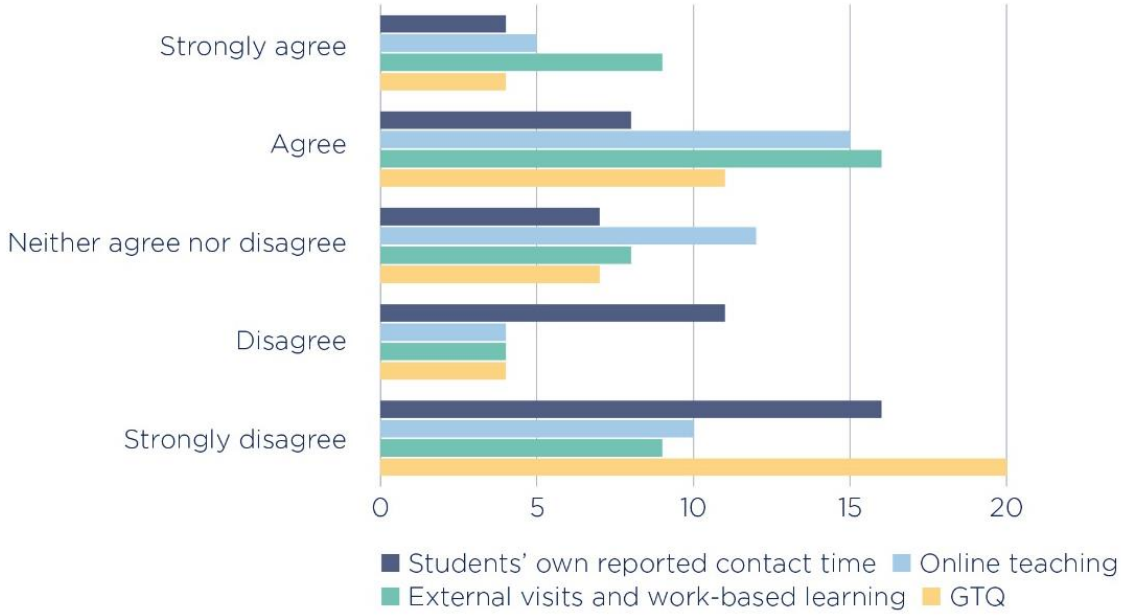
29. A total of 44 providers responded to the survey questions asking whether they used teaching intensity data in their submissions. Of the 44, 20 did refer to provider-declaration data in at least one of their submissions, while 17 of 44 used student survey data. Further education colleges were more likely than not to refer to teaching intensity data (seven of the nine further education college respondents did) and alternative providers were likely not to, while higher education institutions were quite evenly split. Reasons for this relatively low uptake included difficulty interpreting the metrics and putting them into context ('it came too late and [it was] unclear what the number indicated in terms of good or bad', 'provider data was not really referenced as [we] found it difficult to understand where subjects stood in comparisons' and 'we struggled to make head or tail of them to be honest'). Some providers commented that they had been reluctant to use teaching intensity data given the concerns about the reliability of the data, e.g. the low survey response rates and degree of estimation in the provider declarations.

30. After providers had made their submissions, the OfS supplied them with additional information about teaching intensity and included questions in the survey about whether these different data presentations would have been useful to providers if they had had them when writing their submissions[8]. That was to test whether there was anything that we could have done differently that would have improved the usefulness of the metric. In all cases the response was more disagreement than agreement that the additional data would have been useful. One provider observed:

> 'Additional data provided for teaching intensity is not useful if the method for collecting the data and the meaning of the data is fundamentally flawed and opaque. More information is not necessarily better; it would just provide more confusing information'.

31. Providers were asked: 'If this or a similar measure of contact time were to be used in the TEF, to what extent do you agree that it is important to capture GTQ, external visits and work-based learning, online teaching, and students' own reported contact time?'

Figure 1: To what extent do you agree that it is important to capture GTQ, external visits and work-based learning, online teaching, and students' own reported contact time?



32. It appears that there is support for capturing online teaching and external visits and work-based learning, but less support for GTQ and students' own reported contact time. One possible way of interpreting this is that although there is a lack of support for teaching intensity generally, there is a feeling that if teaching intensity data is collected it should not be confined to the GTQ and student perceptions of contact time. It should be noted that the student survey collected

[8] Provider declaration data at a subject-level including medians, the data broken down to course level, the data weighted by actual staff-student ratios rather than weighted in bands (see paragraph 41), unweighted contact time and additional sector-level data.

data on both student perceptions of contact time and student satisfaction with contact time. Only the latter was used in the teaching intensity metric (see paragraph 19).

33. The main reason given for the lack of support for capturing GTQ is that teaching intensity, with its focus on counting hours of traditional face-to-face classroom teaching, was likely to discourage pedagogic innovation. This disagreement would tie in with the support noted in the previous paragraph for capturing online teaching and external visits and work-based learning. It was also frequently commented that teaching intensity measured quantity of teaching rather than quality, that the metrics risked misleading students, and that the measures were highly gameable. One provider commented:

> 'The GTQ or variation is a nonsense. It provides the false security of a "formula" but fails to capture the essence of learning at a university. This could seriously mislead students about the nature of "higher" education in which they are independent adults with choices'.

Panel feedback

34. As teaching intensity was trialled in only five subjects, five of the seven subject panels encountered it in the metrics workbooks (business and law, creative arts and design, engineering and technology, humanities and medical and health sciences). All five were unanimous that the teaching intensity metrics were difficult to interpret and not helpful in assessments. Two panels also commented that teaching intensity, as an input measure, had no place in TEF, which is a student outcomes framework.

> 'the panel had no faith in the teaching intensity data and made a unanimous decision early on in the process that it was simply unusable'
>
> – Humanities Panel report

> 'Teaching intensity was considered to be of no value. There was only one instance where the teaching intensity data served to support a submission that talked about the subject's emphasis on work-based learning'
>
> – Arts Panel report

Analysis

Summarising the provider declaration data into a metric

35. The subject-level metric calculated from the data supplied in the provider declaration for each contact time measure, whether the GTQ, external visits and work-based learning or online teaching, was calculated by considering the distribution of contact time across all students reported on the relevant courses in that subject. The distribution of all of the students on all courses within a subject was considered in order to derive the mean, lower and upper quartile reported in the metric.

36. For example, if a subject had 32 students in total they were separated across courses in the following way:

- course 1 has 20 students with a GTQ of 10

- course 2 has 7 students with a GTQ of 10
- course 3 has 5 students has a GTQ of 70.

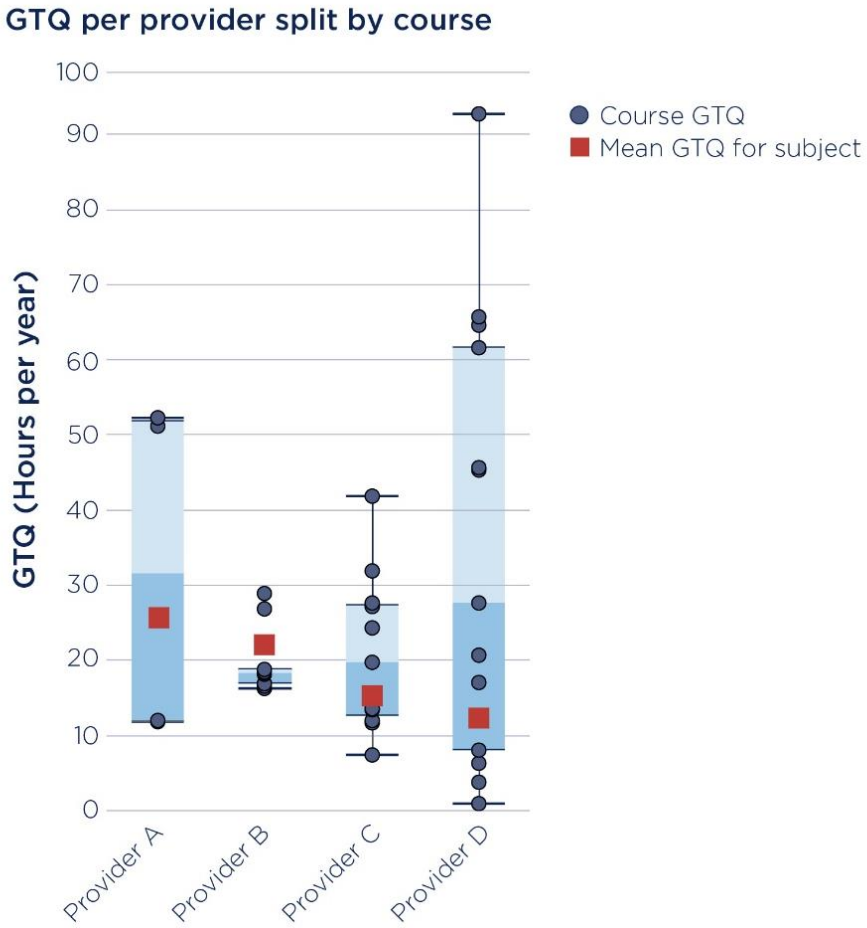
The mean GTQ in this example would be calculated by ((20*10) + (7*10) + (5*70))/32 = 19.4.

37. The lower and upper quartile would consider the 25th and 75th percentile of all students in that subject, and both would have a value of 10 GTQ. This is because the five students with 70 GTQ fall above the 75th percentile.

38. There are cases (such as this example) where the mean of the contact time measure exceeds the upper quartile. This indicates outliers where there are students on courses with significantly higher amounts of contact time than the majority of students on other courses in the subject area. OfS analysts investigated these cases in discussions with providers, and the majority were identified as genuine outliers, often caused by small courses with a large amount of contact time.

39. The box and whisker plot in Figure 2 shows the actual distribution of course GTQ of first-year students in one subject across four different providers. Similar patterns are observed across the majority of providers in all subjects and years of programme of study. In the plot, the providers are ordered from left to right by descending values of mean GTQ. Each blue circle represents the GTQ of a course in that subject. The red square represents the mean GTQ of the distribution. In only one case is the mean GTQ by subject outside of the box and whisker plot, but there is clearly quite a variation in the GTQ of each course across providers within the subject.

Figure 2: Distribution of course GTQ of first year students in one subject across four different providers



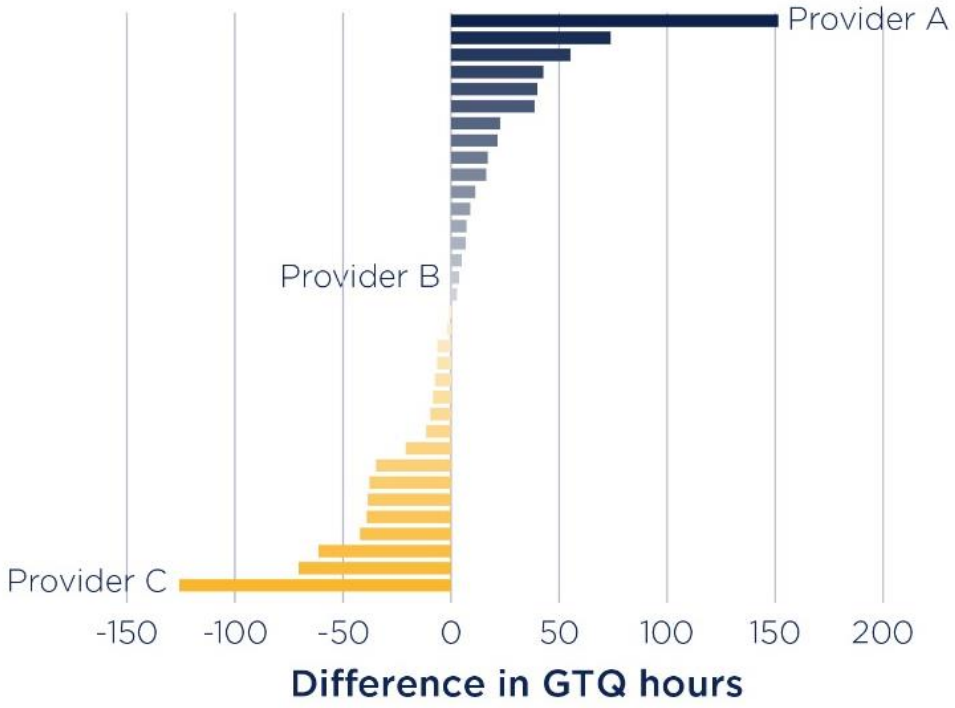
40. As GTQ is a measure which takes into account both contact hours and class sizes, it is possible for a course to achieve a high GTQ in two different ways: either by a high number of contact hours or a lower number of contact hours delivered in classes with a smaller number of students per staff member. GTQ is not a proxy for number of contact hours. Analysis comparing GTQ with total contact hours as reported in the provider declaration found high GTQ scores for courses with both low and high numbers of contact hours.

41. As was outlined in the pilot specification, the GTQ part of the metric was constructed using a method of weighting contact hours based on the student-staff ratio returned in bands rather than the actual student-staff ratio. The weightings that were used per student-staff ratio banding are shown in Table 2. In the provider declaration, 34 of 46 providers also returned the actual student and staff numbers for each contact hour, which meant that a true student-staff weighting could be calculated and compared with the banded rating. Figure 3 represents actual data but anonymises the providers. The values on the x-axis represent the difference in hours per year between the GTQ calculated by using weightings associated with the student-staff ratio bandings and the GTQ calculated on the basis of actual student-staff ratio. A provider perfectly represented by the banded weightings would appear at 0 (similar to Provider B); this clearly applied to very few of the providers in the sample, and in some cases the positive (Provider A) and negative (Provider C) discrepancies caused by the banding were severe.

Table 2: Student-staff ratio band weightings

Student-staff ratio (X)	Weighting
X ≤ 2	2/3
2 < X ≤ 8	1/5
8 < X ≤ 20	1/14
20 < X ≤ 40	1/30
40 < X	1/75

Figure 3: Provider representation of the effect of the use of banded weightings



42. Overall the specified approach to creating the provider declaration of teaching intensity resulted in metrics that were not only difficult for providers and panel members to interpret, but in many cases not very representative of provision. Moreover, as noted at paragraph 22, it is likely that there were assumptions and inaccuracies in providers' reported data.

Student survey

43. The student survey asked four student experience questions as shown at paragraph 15 above and it was found that there was a highly positive correlation between students' responses to these questions. Principal component analysis, a statistical technique which estimates the number of concepts underlying the questions in a survey, suggested that there was only one

concept underlying the four questions. On this basis, the four questions were converted into a single scale when released to providers and in all subsequent analysis.

44. The OfS was able to map to HESA or ILR data for just over 2,580 of the 5,650 full and partial survey responses received. Of those not mapped, just under 2,370 of the 3,070 were first-year students. The purpose of linking to HESA or ILR data was to give information about a student for the purpose of benchmarking. The low response rates and the consequently low number of survey responses mapped to HESA or ILR data meant that there was too much data sparsity across benchmarking factors to be able to produce a robust benchmark. If benchmarking had been possible, the factors that we would have used would have followed the approach made in benchmarking the NSS scales in the TEF metrics. When looking at the representativeness of the benchmarking combinations, 28 of a possible 83 combinations were represented in the student survey results. Overrepresentation was identified in three categories:

    - business and management in both full and part-time, young and mature, white, degree level students with no disability
    - engineering in both full and part-time, young, white, degree or other undergraduate level students with no disability
    - history and archaeology in full-time, young, white, degree level students with no disability.

45. The OfS additionally looked to understand where there was any bias in the partial responses. Looking separately by subject, by year, and by subject and year, we found that no combination had a higher proportion of partial responses compared with the equivalent proportions of expected responses overall and hence found no bias in the partial responses.

Correlation analysis

46. The OfS analysts looked at the relationship between the GTQ, the overall amount of contact time (not weighted on the basis of student-staff ratio bandings), the levels of satisfaction and the self-reported number of contact hours from the student survey of teaching intensity by subject and year of programme of study.

47. The results of looking at these relationships are listed below. In each case, analysts assessed the goodness of fit, using the associated R-squared value. Cases were only considered where a provider has given information on contact time and has had more than 10 student survey responses in the categories concerned.

Relationship between GTQ and student satisfaction

48. No correlation was found between the GTQ and the satisfaction from the student survey, looking:

    - by subject and year
    - by aggregating all years of programme of study together within each subject
    - by aggregating all subjects within each year of programme of study
    - by aggregating all subjects and years together.

Student perceptions of contact hours

49. The overall amount of contact time reported by students is taken from the student survey responses. The survey's primary focus was to ask questions about student satisfaction with the contact time provided, but students were additionally asked to quantify how many hours per week on average they spent in teaching and learning activities. The survey questions used for this measure were: 'Thinking back over this term, how many hours per week on average were scheduled for face-to-face teaching such as lectures, seminars, tutorials, project supervision, demonstrations, practical classes and workshops, or supervised time in a studio or workshop?' and 'Thinking back over this term, how many hours per week on average were scheduled for online learning activities facilitated by a tutor, such as live lectures, question and answer sessions, discussion forums, group or individual presentations and workshops?' These teaching and learning activities were akin to those specified for the provider declaration in the provider guidance, although necessarily less detailed and less defined because of the limitations of space in the survey.

50. The primary reason for asking these questions, as recommended by the cognitive testing of the survey in its development phase, was to encourage students to think about their contact time before responding about satisfaction, but students' answers to these questions also enabled the OfS to compare the differences between perceptions of contact time reported by the students and the contact time reported by the provider. To make students' responses to these questions comparable to the time reported by the provider, students' reported hours were multiplied by 30, assuming a 30-week year of programme of study. In addition we have had to assume that the average contact time offered in the most recent term is consistent across all terms throughout the year. It should also be noted that the students responding to the survey in 2017-18 were not the same cohort as those for whom the provider declaration was completed (usually based on the 2015-16 HESA return) and, as noted in paragraph 17, students in scope for the survey were defined more broadly than students in scope for the provider declaration. In short there are differences between the student survey population and the provider declaration population that most likely account for some of the divergence between the two.

51. There is a high amount of variation in the student perceptions of contact time per provider, subject of study and year of programme. This may be explained by the different structure of programmes within a provider's subject and the differences outlined in paragraph 50, but there are a number of combinations that have a difference of more than 15 hours of contact time per week when looking at the difference between the upper and lower quartiles at this level.

52. We have looked at the relationship between the overall amount of contact time reported by the provider and that reported by the student, looking by subject and year, by aggregating all years of programme of study together within each subject, by aggregating all subjects within each year of programme of study and by aggregating all subjects and years. In each of these cases, we have compared a provider's mean amount of contact time against the mean amount of contact time reported by the student per by group, and have also done equivalent comparisons using the median. We have found no correlation in any of these cases. Generally, the trend lines of the relationship appear nearly horizontal, which confirms that irrespective of the assumption made on the number of weeks taught per year of programme of study this has not caused the lack of correlation.

Conclusion

53. Analysis suggests that whatever is measured in the teaching intensity metric does not strongly correlate either with students' satisfaction with the contact time they receive or with providers' reports of the hours provided. In addition, providers and panel members alike found the metrics difficult to understand and of very little use in the assessment process, and cost survey data suggests that collecting teaching intensity data would substantially increase the cost and burden of the exercise for the sector. Contact hours information is known to be an issue that is important to students and applicants[9], but adopting teaching intensity as a TEF metric does not appear to be an effective approach to conveying this information in a way meaningful to applicants or helpful in informing student choice.

[9] See for example the recent research commissioned by the OfS and led by a consortium of students' unions, which found that 91 per cent of students surveyed considered the number of contact hours per week 'very' or 'somewhat' important. 'Value for money: The student perspective', https://studentsunionresearch.com, page 16.