Data modelling to inform proposals for a revised approach to assessing and rating quality

Benchmarking factors

Introduction

1. We use benchmarking in the Teaching Excellence Framework (TEF) to take account of the mix of courses and students at a provider and to indicate how well that provider has performed compared with performance for similar types of students on similar types of courses across the higher education sector as a whole. Similar types of students and courses are identified using **benchmarking factors**.

2. The benchmark for a provider is calculated as a weighted average which represents the outcomes that would have been achieved by the provider if it retained its mix of students and courses, but its outcomes across the benchmarking factors were replaced by the rates for those student groups across the sector as a whole.

3. We then compare the provider's actual value against its benchmark value, to understand how well the provider has performed for its particular mix of students and courses. For example, if a provider has a continuation rate of 90 per cent, and its benchmark value is 85 per cent, it has a higher continuation rate than similar students on similar courses, across the sector as a whole.

4. The factors we use in calculating the benchmark values relate to individual students at a provider and vary according to the measure we are benchmarking. We use benchmarking factors that, across the sector as a whole, are most correlated with the outcomes and experiences we are measuring once other factors have been controlled for, unless we consider it undesirable to control for them from a policy perspective. These factors relate to characteristics of courses (such as subjects and level of study) and students or graduates (such as their age and other personal characteristics, and whereabouts in the country they are working). Each factor is made up of categories. For instance, the categories for the 'level of study' factor include 'first degree' and 'other undergraduate'.

5. This section documents the benchmarking factors we propose to use for the student experience, use of skills and earnings measures, and the analysis and considerations that led us to these proposals. We are not proposing any changes to the benchmarking factors for the continuation, completion or progression measures.

6. In developing these proposals, we have followed our guiding principles for the selection and application of benchmarking factors.[1] We have also drawn on previous research and consultation carried out to inform the current student outcomes and TEF indicators,[2] and the

[1] See Guiding principles for benchmarking factors - Office for Students.

[2] See Review of the selection and grouping of benchmarking factors - Office for Students.

review of benchmarking factors for the National Student Survey (NSS) in 2023.[3] In all cases, the analysis is based on a population of undergraduate students registered at OfS-registered providers.

Review of benchmarking factors for student experience measures

7. We have recently developed two new NSS 'Association between characteristics of students' (ABCS) measures (one for full-time students and one for part-time students), to consider whether they should be used for benchmarking the NSS indicators. These measures are explained in detail in the section below entitled 'ABCS measures for NSS'.

8. We reviewed existing NSS benchmarking factors to consider whether we should use the new NSS ABCS, either as an additional factor or in place of current student characteristics. The potential advantage of using ABCS in benchmarking is that we can consider a wide range of student characteristics, and combinations of these characteristics, in a single factor. This would allow us to take account of variation in students' responses based on their characteristics without having as many benchmarking factor groups and groups with small numbers of students. As explained in the 'Benchmarking factors' section of Proposal 8 in the consultation document, this has the potential to increase the effectiveness of our benchmarking approach by reducing levels of self-benchmarking. Using the NSS ABCS, instead of several separate student characteristics, would be consistent with our current approach for student outcomes benchmarking factors.

Current and proposed NSS benchmarking factors

9. Table 1 shows the current benchmarking factors we use for student experience measures. These differ for full-time, part-time and apprenticeship students.

Table 1: Current benchmarking factors

Benchmarking factor	Student-experience full-time	Student experience part-time	Student experience apprenticeship
Year of survey	✓	✓	✓
Level of study (First degree, Other undergraduate, Undergraduate with postgraduate components)	✓	✓	✓
Subject of study	✓ (34 CAH level 2 groups)	✓ (10 Broad subject groups)	✓ (10 Broad subject groups)
Age on entry (Under 21 or unknown, 21 to 30, 31 and over)	✓	✓	✓
Disability (Disability reported, no disability reported)	✓	✓	✓

[3] See OfS, Benchmarking in the NSS publication [PDF].

Benchmarking factor	Student-experience full-time	Student experience part-time	Student experience apprenticeship
Ethnicity (Asian, Black, Mixed, Other, Unknown or White, non-UK domiciled)	✓	✓	✓
Sex (Female or Other, Male	✓	x	x
Total possible distinct benchmarking groups	29,376	4,320	4,320

10. Table 2 shows the benchmarking factors we propose to use following the review.

Table 2: Proposed benchmarking factors

Benchmarking factor	Student experience full-time	Student experience part-time	Student experience apprenticeship
Year of survey	✓	✓	✓
Level of study (First degree, Other undergraduate, Undergraduate with postgraduate components)	✓	✓	✓
Subject of study	✓ (34 CAH level 2 groups)	✓ (10 Broad subject groups)	✓ (10 Broad subject groups)
NSS ABCS (NSS ABCS quintiles 1 to 5 for the relevant mode of study; Non-UK domiciled)	✓	✓	✓
Total possible distinct benchmarking groups	2,448	720	720

Method for review

11. For this review, our evaluation of the potential benchmarking factors involved two steps. In step one we carried out statistical modelling following the same approach as the January 2022 evaluation of benchmarking factors.[4] The aim of this step was to determine whether the results of statistical modelling based on current factors remain consistent with the previous review and whether use of the new NSS ABCS measures can allow us to make improvements to the model.

[4] See Review of the selection and grouping of benchmarking factors - Office for Students.

12. We produced separate models for each mode of study (full-time and part-time). Previous analysis has shown that the benchmarking factor groups can become quite small when looking at the apprenticeship mode of study and, as a result, it can be difficult to obtain modelling results for this mode. We therefore propose to continue with our current approach of applying the same benchmarking factors to apprenticeship students as we do to part-time students.[5] This approach is primarily driven by the smaller size of the apprenticeship sector, as including all the factors used for full-time students would mean we were too often making comparisons between very small groups of students.

13. The aim of step two was to determine how the proposed benchmarking factors and groupings affected the statistical integrity of the benchmarking approach. This step primarily considered the contribution to its own benchmark made by each provider.

14. These steps help us decide which factors to prioritise. In prioritising factors, we seek to balance retaining factors with higher levels of explanatory power with minimising the number of students likely to be in small groups and the number of providers with substantial contributions to their own benchmarks.

15. We produce and benchmark student experience measures for each NSS theme. For simplicity, we currently have a single set of benchmarking factors across all themes and propose to continue with this approach. This is supported by previous analysis of NSS themes, which has shown consistency in the factors that are correlated across the NSS questions and themes. Due to the consistency we know to exist across questions, for brevity, the analysis presented here focuses on the results from modelling on Question 16: 'How well have teaching staff support your learning?'. This question has been chosen as levels of positivity for this question have been shown to be highly correlated with all other questions across the survey, and more consistently so than any other question.

Review of benchmarking factors for full-time students

Step one

16. In step one, we produced a statistical model to reflect the three options under consideration:

    - Option 1 is based on current benchmarking factors (as set out in Table 1)
    - Option 2 additionally includes the new ABCS measure as a factor
    - Option 3 uses the new ABCS measure in place of other student characteristics that are already reflected in the NSS ABCS measure.

17. We included entry qualifications in modelling in previous NSS benchmark reviews but found it not to be significant. As we do not have any evidence to suggest there would be a change in its importance, we have not reconsidered this factor.

18. Table 3 summarises the results of the statistical model for the full-time mode of study and shows the maximum size of the estimated differences (for the categories with the largest

[5] Apprenticeship students have not been included when constructing the ABCS measures, but ABCS quintiles have been assigned to them using the NSS ABCS full-time measure. We consider this the best match for apprenticeship students considering the student experiences of these students.

significant difference in each case) for each of the shortlisted factors. It aims to provide an indication of the relative correlation with the outcome measured.

19. To give an example of how to interpret this type of table, the top row of the table shows 3.5 for age on entry for Option 1. To calculate this, we model the effect of each age group when combined with all the other factors in the model. Then we compare the age groups with largest and the smallest effects. If the 'Under 21' group had an estimated effect of -1 and the 'Over 31' group had an effect of 2.5, the difference between those gives 3.5. The larger this difference, the more of the variation between different students is explained by the factor, and the more difference it makes when included in the model. These effects change when other factors are included or removed – in Option 2 we can see that there is less difference between the effects of the age groups, because some of the variation is explained by the added ABCS factor.

Table 3: Maximum estimated differences for the shortlisted factors (Question 16 Full-time)

Factor	Option 1: Current factors	Option 2: Current factors plus ABCS	Option 3: ABCS in place of student characteristics
Age on entry	3.5	2.1	N/A
Disability	3.2	1.9	N/A
Ethnicity	3.9	2.1	N/A
Level of study	2.9	3.1	2.6
Sex	1.1	2.1	N/A
Subject	6.0	6.0	5.4
Survey year	6.5	6.4	6.4
NSS ABCS	N/A	4.6	6.1

20. The results of statistical modelling of current factors remain consistent with previous analysis, as the current factors demonstrate strong correlation with the positivity measures. The modelling results also suggest that the NSS ABCS measure captures much of the variation previously explained by individual student characteristics. This supports replacing several separate characteristics with a single NSS ABCS measure, as it would simplify the benchmarking approach while maintaining explanatory power.

Step two

21. To understand how the proposed benchmarking factors and the resulting benchmarking groups would impact the benchmarking process, we calculated the number of distinct benchmarking groups and the contribution of providers towards their own benchmarks. Tables 1 and 2 at the start of this section show that the number of possible distinct benchmarking groups drops from nearly 30,000 with the current approach (Option 1) to less than 2,500 (Option 3) if the proposed factors are used. Option 2 would have almost 150,000 possible groups.

22. The statistic showing contribution of providers towards their own benchmarks indicates the influence of the provider's own students on the sector averages that inform the calculation of the provider's benchmark, and helps us understand the risk of self-benchmarking. The risk of self-benchmarking arises when benchmarking groups are defined at such a detailed level that only very small numbers of students possess each unique combination of the student and course characteristics that we have selected to act as benchmarking factors. In such a scenario, the provider's own students would be making a substantial contribution to the calculation of its benchmark, making the calculation less robust and the resulting benchmark value less meaningful. We remain of the view that it is essential that this risk is minimised as far as possible. However, we are also aware that the diversity of the higher education sector means that we cannot mitigate this risk entirely and our proposed benchmarking factors tolerate a risk of self-benchmarking on a small scale.

23. Table 4 shows the proportion of providers that contribute more than 5 per cent, 20 per cent, or 50 per cent towards their own benchmarks, based on the modelled options.

Table 4: Providers' contributions to their own benchmarks based on proposed benchmark factors (Question 16 Full-time)

		Percentage of providers with at least (X%) contribution to their own benchmark		
Model option	**Number of providers**	**>5%**	**>20%**	**>50%**
Option 1	351	52%	9%	1%
Option 2	351	79%	19%	2%
Option 3	351	20%	3%	0%

24. Under Option 2 (including the NSS ABCS as an additional factor), the number of unique benchmarking groups would increase and the risk of self-benchmarking would also increase. Replacing the student characteristics of age, ethnicity, disability and sex with NSS ABCS, however, reduces the potential for self-benchmarking in comparison to current factors. We consider that Option 3 achieves the most appropriate balance of policy objectives and statistical properties of the three options and propose it is therefore consistent with the principles for the selection and application of benchmarking factors.

Review of benchmarking factors for part-time students

Step one

25. We repeated the same steps for part-time students. We did not consider entry qualifications or sex for part-time, as these were found to be not significant in previous NSS benchmark reviews, and we do not have any evidence to suggest there is change in their importance.

26. Table 5 summarises the results of the statistical model for the part-time mode of study and shows the maximum size of the estimated differences (for the categories with the largest significant difference in each case) for each of the shortlisted factors.

Table 5: Maximum estimated differences for the shortlisted factors: Question 16, Part-time

Key: * means not all categories within the factor had statistically significant estimate differences. We have not included insignificant categories when calculating the maximum estimated difference.

Factor	Option 1: Current factors	Option 2: Current factors plus ABCS	Option 3: ABCS in replace of characteristics
Age on entry	2.1	0.2*	N/A
Disability	4.0	1.9	N/A
Ethnicity	4.2	2.1*	N/A
Level of study	1.4	1.2	1.1
Subject	5.8	5.8	5.3
Survey year	2.9	2.9	2.7
NSS ABCS	N/A	6.9	8.4

27. As for the full-time mode, the results of statistical modelling of current factors remain consistent with previous analysis and the current factors demonstrate strong correlation with the positivity measures. The new NSS ABCS measure has a stronger correlation than each of the student characteristics individually. As it has taken account of the separate student characteristics in its preparation, this supports replacing the separate characteristics with a single NSS ABCS measure for part-time also.

Step two

28. To understand how the proposed benchmarking factors and the resulting benchmarking groups would impact the benchmarking process, we calculated the number of distinct benchmarking groups and the contribution of providers towards their own benchmarks. Tables 1 and 2 at the start of this section show that the number of possible distinct benchmarking groups drops from nearly over 4,000 with the current approach (Option 1) to less than 750 if the proposed factors are used (Option 3). Option 2 would have over 20,000 possible groups.

29. Table 6 shows the proportion of providers that contribute more than 5 per cent, 20 per cent, or 50 per cent towards their own benchmarks under each of the three options.

Table 6: Provider's contributions to their own benchmarks based on proposed benchmark factors (Question 16, Part-time)

		Percentage of providers with at least (X%) contribution to their own benchmark		
Model option	Number of providers	>5%	>20%	>50%
Option 1	208	66%	19%	5%
Option 2	208	88%	44%	11%
Option 3	208	34%	8%	2%

30. Again, as for the full-time mode, under Option 2 the number of unique benchmarking groups would increase and the risk of self-benchmarking would also increase. Option 3 reduces the potential for self-benchmarking in comparison to the current set of factors. As the findings for the part-time mode are consistent with those for full-time, they support our proposed approach of replacing individual student characteristics with NSS ABCS.

Application of benchmarking factors to NSS themes

31. We discussed in paragraph 15 that we had conducted our analysis on Question 16 for simplicity. In order to test whether these findings still stand at theme level, we have looked at one of the NSS themes, 'Teaching on my course', for full-time students to see if the statistical modelling provides findings consistent with that of Question 16. This theme was selected to use as it is the first theme in the survey and does not contain Question 16.

Table 7: Maximum estimated differences for the shortlisted factors – Teaching on my course theme (Full-time)

Factor	Option 1: Current factors (Question 16)	Option 1: Current factors (Teaching theme)	Option 3: ABCS in replace of characteristics (Question 16)	Option 3: ABCS in replace of characteristics (Teaching theme)
Age on entry	3.5	5.8	N/A	N/A
Disability	3.2	1.8	N/A	N/A
Ethnicity	3.9	2.8	N/A	N/A
Level of study	2.9	3.2	2.6	3.1
Sex	1.1	0.8	N/A	N/A
Subject	6.0	8.0	5.4	7.9
Survey year	6.5	3.6	6.4	3.6
NSS ABCS	N/A	N/A	6.1	4.8

32. The model outputs show that, for the 'Teaching on my course' theme, the proposed factors are still significant – with subject, survey year and level having the strongest explanatory power among the factors we are proposing (Option 3).

33. We can see that there are differences in the effects between this theme and Question 16 when we apply the proposed factors; in particular, age is currently more significant for this theme than the proposed ABCS factor would be. We expect differences like this between different questions and themes, and they are part of the trade-off we make when selecting a single set of benchmarking factors for all themes rather than different factors for each theme. However, overall, we can see that all the factors in our chosen model have significant large effects on the 'Teaching on my course' theme, and the other factors did not increase in effect size between Options 1 and 3. We therefore consider that our proposed factors will result in an effective benchmarking approach across themes.

Analysis of benchmarking factors for the 'use of skills' measure

Introduction and proposed factors

34. The 'use of skills' measure is a new measure based on a graduate reflection question from the Graduate Outcomes survey: 'I am utilising what I learnt during my studies in my current activity/study/work'. Further details about this measure can be found in the consultation document (Proposal 14 and Annex E: Definitions of new post-study measures).[6] In Part 4.9 of the draft guidance,[7] and below, we set out the proposed benchmarking factors for the new measure. In this section we set out the analysis and considerations that led us to these proposals.

35. Table 8 shows the benchmarking factors we propose to use following the analysis. We consider that the proposed factors strike an appropriate balance between the policy objectives of benchmarking and the need to maintain statistical robustness. In prioritising factors, we have sought to balance retaining factors with higher levels of explanatory power with minimising the number of students likely to be in small groups and, consequently, the number of providers with substantial contributions to their own benchmarks.

Table 8: Proposed benchmarking factors for the use of skills measure

Benchmarking factor	Skills utilisation: full-time	Skills utilisation: part-time	Skills utilisation: apprenticeship
Level of study	✓	✓	✓
Subject of study	✓ (34 CAH level 2 groups)	✓ (10 Broadly defined subject groups)	✓ (10 Broadly defined subject groups)
Entry qualifications	✓ (10 groupings)	✓ (5 groupings)	✓ (5 groupings)
ABCS progression quintile (quintile 1, quintiles 2 and 3, quintiles 4 and 5, and unknown)	✓	✓	✓
Geography of employment quintile (quintile 1, quintiles 2 and 3, quintiles 4 and 5, and unknown)	✓	✓	✓
Year qualification obtained	✓	✓	✓

36. The remainder of this section outlines the approach and analysis which led us to propose this set of benchmarking factors.

Candidate benchmarking factors

37. In previous benchmarking reviews, we considered a potential longlist of benchmarking factors and then narrowed these down to a shortlist based on the suitability of each factor from a policy perspective and the availability of data. As this shortlisting process had already been

[6] See Proposals for a revised approach to assessing and rating quality - Office for Students.

[7] Available at Annexes and supporting documents - Office for Students.

carried out for the progression measure, we started our analysis for the use of skills measure by considering those shortlisted factors, as we considered them most relevant. The factors we assessed for the use of skills measure are:

- Subject (CAH2 level)
- Entry qualifications
- Level of study
- Year qualification obtained
- Geography of employment quintile[8]
- Age
- Disability type
- Ethnicity
- Sex
- IMD quintile
- ABCS progression quintile.

38. Rather than creating new ABCS measures specifically for use of skills, we have used the existing ABCS progression measures. These model the likelihood of being in highly skilled employment or further study when surveyed, based on the characteristics held by the student. Our analysis shows that there is enough of an overlap between progression outcomes and use of skills outcomes to suggest that progression ABCS is useful for accounting for some of the differences between students which impact their likelihood of using what they learned in higher education. Therefore, we consider it is better to use this as a factor than to increase complexity by creating another similar ABCS measure. The ABCS is used in the same way as in progression, with unknown values being mapped to ABCS = 5.

39. We have carried out statistical modelling using the same approach as for the review of benchmarking factors for student experience measures, which is described in the subsection 'method for review' above. Again, we ran models separately for the full-time and part-time modes of study, and we propose to apply the same benchmarking factors to apprenticeship students as we do to part-time students.

Analysis of benchmarking factors for full-time students

40. As explained in the 'Benchmarking factors' section of Proposal 8 in the consultation document, to maintain statistical integrity of the benchmarking process we seek to avoid creating benchmarking groups defined at a very granular level which are likely to each contain very small numbers of students. It is therefore not desirable to use all of the factors we assessed, and we need to prioritise these. We have modelled two options for consideration:

- Option 1 uses the same set of benchmarking factors as used for the progression measure

[8] See A geography of employment - Office for Students.

- Option 2 uses most of the same factors as Option 1 but includes the most significant student characteristics instead of the progression ABCS groups.

41. Table 9 summarises the results of the statistical model for the full-time mode of study. The columns show the maximum size of the estimated differences (for the categories with the largest significant difference in each case) for each of the shortlisted factors. They aim to provide an indication of the relative correlation with the outcome of the use of skills measure.

42. The table only shows estimated differences that are statistically significant at the 95 per cent level and excludes 'unknown' and similar categories. In this case, all of the factors we assessed showed statistically significant estimated differences for all categories. Note that many factors had multiple differences, and we considered differences across all of the categories, as well as the numbers of students associated with each category, when making our judgements.

Table 9: Maximum estimated differences for the shortlisted factors (Full-time)

Factor	Full-time (all factors)	Option 1 (same factors as the progression measure)	Option 2 (using Sex and Age, dropping ABCS)
Age on entry	5.9	N/A	6.3
Progression ABCS quintile	2.7	4.2	N/A
Deprivation quintile (IMD)	1.0	N/A	N/A
Disability	1.7	N/A	N/A
Entry qualifications	8.2	9.1	9.5
Ethnicity	1.7	N/A	N/A
Geography of employment quintiles	3.6	3.5	3.5
Level of study	8.3	8.2	8.2
Sex	2.1	N/A	2.0
Subject	26.4	29.0	26.4
Qualifying year	1.1	1.2	1.2

43. From the data presented in Table 9, we can see that **subject**, **entry qualifications**, and **level of study** are the factors associated with the largest estimated differences and the strongest explanatory power. We consider that we should therefore prioritise these factors and have included them in all the models. **Geography of employment quintile** shows the next strongest association with use of skills outcomes after age, so we have also included that in all the models.

44. We also consider **qualifying year** to be an important factor to prioritise, despite it not having one of the strongest levels of explanatory power. Including qualifying year ensures consistency with the factors used in the progression and student experience measures. It also allows the model to account for events that may have affected outcomes in specific years. For example, some years of data may continue to be influenced by the impacts of the COVID-19 pandemic. It also supports comparability for providers that are assessed in different years, given the decision to move to a rolling cycle of assessments.

45. We can see that either the **ABCS group** (in Option 1) or **Sex and Age** (in Option 2) are significant.

46. On balance we consider that Option 1 is preferable to Option 2, for the following reasons:

    a. Consistency with the progression benchmarking factors would make the approach easier to understand, and

    b. The ABCS group takes into account a wider range of student characteristics, and combinations of those characteristics. Option 2 does not, for example, take account of ethnicity or disability.

47. We also need to check that Option 1 will not have an adverse effect on the benchmarking process. To check this, we examined the number of benchmarking groups generated, the proportion of students in very small groups, and the extent to which providers contribute to their own benchmarks. As explained in the 'Benchmarking factors' section of Proposal 8 in the consultation document, where a provider's own students make a substantial contribution to the benchmark calculation this limits the value of the benchmark and we should therefore seek to minimise the risk of self-benchmarking.

Table 10: Number of categories, combinations, and small groups from the proposed factors

Factor	Skills utilisation (Full-time)
ABCS progression	6
Entry qualifications	10
Geography of employment	3
Level of study	3
Subject (CAH2)	34
Year of indicator	4
Summary	
Number of possible combinations	73,440
Number of populated combinations	24,984
% of students in very small groups (1 to 5)	3.5%
% of students in small groups (1 to 20)	19.5%

Table 11: Providers' contributions to their own benchmarks using proposed factors

Percentage of providers with at least (X%) contribution to their own benchmark	Skills utilisation (Full-time)			
	Number of providers	>5%	>20%	>50%
All undergraduates	326	92%	7%	1%

48. For the set of factors proposed in Option 1, our view is that the number of theoretical and populated benchmarking groups is a manageable number, minimises the number of students in very small groups, and does not result in too much self-benchmarking.

49. In contrast, if we were to include all the personal characteristic factors without the ABCS factor, the number of possible combinations increases to 2,643,840 which would lead to a much higher risk of self-benchmarking.

50. We consider Option 1 achieves an appropriate balance of policy objectives and statistical properties and proposing it is therefore consistent with our principles for the selection and application of benchmarking factors.

Analysis of benchmarking factors for part-time students

51. Having identified Option 1 as our preferred option for full-time, we modelled this for part-time.

52. Table 12 shows the results of the statistical model for the part-time mode of study for all factors and Option 1. It displays the maximum size of the estimated differences (for the categories with the largest significant difference in each case) for each of the candidate factors. It aims to provide an indication of the relative correlation with the outcome of the skills utilisation measure.

53. The table only shows estimated differences that are statistically significant at the 95 percent level and excludes 'unknown' and similar categories. Note that many factors had multiple significant differences, and we considered differences across all of the categories, as well as the numbers of students associated with each category, when making our judgements.

Table 12: Maximum estimated differences for the shortlisted factors (Part-time)

Factor	Part-time (all factors)	Option 1 (using ABCS in place of characteristics)
Age on entry	1.9	N/A
ABCS quintile	2.0	2.0
Deprivation quintile (IMD)	Not significant	N/A
Disability	2.4	N/A
Entry qualifications (5 groups)	1.1	2.2
Ethnicity (2 groups)	1.9	N/A
Geography of employment quintiles	0.6	0.7

Factor	Part-time (all factors)	Option 1 (using ABCS in place of characteristics)
Level of study	1.0	1.1
Sex	1.2	N/A
Subject (broad subject)	14.1	14.2
Year	0.8	0.9

54. For part-time students, some categories have been combined because individual groups were too small for reliable analysis. Table 12 shows that the IMD quintile was not statistically significant, so we deprioritised this factor.

55. From the data in the second column, we can see that **Subject** is the factor with the biggest estimated differences and the strongest explanatory power and we consider this should therefore be prioritised as a benchmark factor. We also consider that **qualifying year** should be prioritised for the same reasons as we set out for full-time.

56. While **geography of employment quintile, level of study, and entry qualifications** have lower explanatory power than some other factors, we consider that they should be prioritised to maintain consistency with the benchmarking factors selected for full-time students, provided their inclusion does not materially increase the risk of self-benchmarking. In particular, the inclusion of geography of employment quintile would allow the benchmark to account, to some extent, for variation in outcomes associated with the area in which graduates live and work, including differences in local labour market opportunities.

57. We can also see that the **ABCS progression quintile** is a significant factor, both when combined with all other factors and in Option 1. To understand the impact of Option 1 on the benchmarking process, we examined the number of benchmarking groups generated, the proportion of students in very small groups, and the extent to which providers contribute to their own benchmarks. As explained in the 'Benchmarking factors' section of Proposal 8 in the consultation document, where a provider's own students make a substantial contribution to the benchmark calculation this limits the value of the benchmark and we should therefore seek to minimise the risk of self-benchmarking.

Table 13: Number of categories, combinations, and small groups from the proposed factors (Option 1)

Factor	Skills utilisation (Part-time)
ABCS progression	6
Entry qualifications	5
Geography of employment	3
Level of study	3
Subject (CAH2)	10
Year of indicator	4

Factor	Skills utilisation (Part-time)
Summary	
Number of possible combinations	10,800
Number of populated combinations	3,334
% of students in very small groups (1 to 5)	6.3%
% of students in small groups (1 to 20)	34.9%

Table 14: Providers' contributions to their own benchmarks using the proposed benchmarking factors (Option 1)

Percentage of providers with at least (X%) contribution to their own benchmark	Skills utilisation (Part-time)			
	Number of providers	>5%	>20%	>50%
Full-time, all undergraduates	175	93%	15%	3%

58. For the set of factors proposed in Option 1 our view is that the number of theoretical and populated benchmarking groups is a manageable number, minimises the number of students in very small groups, and does not result in too much self-benchmarking.

59. If we were to include all the personal characteristic factors without the ABCS factor, the number of possible combinations increases to 43,200 which would lead to a higher risk of self-benchmarking.

60. Taking account of both the modelling results and the benefits of including the intersectional ABCS progression quintile measure, we consider that Option 1 represents an approach that achieves a reasonable balance between maximising explanatory power and limiting the risk of self-benchmarking.

Analysis of benchmarking factors for the earnings measure

Introduction and proposed factors

61. The earnings measure is a new measure calculated using the Longitudinal Education Outcomes (LEO) dataset, which links student records with tax data to determine whether graduates were employed and how much they earned. Further details about this measure can be found in the consultation (Proposal 14 and Annex E: Definitions of new post-study measures). The draft guidance (Part 4.9) sets out the proposed benchmarking factors for this new measure. This section sets out the analysis and considerations that led us to these proposals. Note that the analysis in this section is based on data which excludes the small proportion of graduates who were returned to the Student Alternative Record.

62. Table 15 shows the benchmarking factors we propose to use following the analysis. We consider that the proposed factors strike an appropriate balance between the policy objectives of benchmarking and the need to maintain statistical robustness.

Table 15: Proposed benchmarking factors for the earnings measures

Benchmarking factor	Full-time	Part-time	Apprenticeship
ABCS progression quintile (quintile 1; quintiles 2 and 3; quintiles 4 and 5; unknown)	✓	✓	✓
Entry qualifications	✓ (10 groupings)	✓ (3 groupings)	✓ (3 groupings)
Geography of earnings quintile (quintiles 1 to 5)	✓	✓ (3 groupings)	✓ (3 groupings)
Level of study (first degree; other undergraduate; undergraduate with postgraduate component)	✓	✓	✓
Subject of study (CAH level 2)	✓	✓	✓
Qualifying year	✓	✓	✓

63. The remainder of this section outlines the approach and analysis which led us to propose this set of benchmarking factors.

Candidate benchmarking factors

64. As for the use of skills measure, and for the same reasons, we started by considering the factors that were shortlisted for the progression measure, plus the ABCS progression quintile. However, we replaced the geography of employment quintile with a newly developed geography of earnings quintile.[9] The factors we have assessed for the earnings measure are therefore:

- ABCS progression quintile
- Age on entry
- Deprivation quintile (IMD)
- Disability
- Entry qualifications
- Ethnicity
- Geography of earnings quintile
- Level of study
- Sex
- Subject of study (level 2 of the Common Aggregation Hierarchy)
- Year.

65. The geography of earnings quintile system assigns travel to work areas (TTWAs) to five groups from 1 to 5. Each quintile represents a fifth of the population used in the analysis. The measure used to put areas into groups is the same as the earnings measure being proposed

[9] See Geography of earnings, available at A geography of employment and earnings - Office for Students.

in this consultation. Areas with the highest earnings measure are in quintile 5, and areas with the lowest earnings measure are in quintile 1.

66. We have carried out statistical modelling using the same approach as for the review of benchmarking factors for student experience measures, which is described in the subsection 'method for review' above. Again, we ran models separately for the full-time and part-time modes of study, and we propose to apply the same benchmarking factors to apprenticeship students as we do to part-time students.

67. We have modelled two options for consideration for the earnings measure:

- Option 1 uses the same set of benchmarking factors as used for the progression measure, but with geography of earnings instead of the geography of employment
- Option 2 uses most of the same factors as Option 1 but includes all significant student characteristics instead of the progression ABCS groups.

Analysis of benchmarking factors for full-time students

68. Table 16 shows the largest significant estimated differences for each of the shortlisted factors. It aims to provide an indication of the relative correlation with the outcome of the earnings measure. When compiling the table, we only compared estimated differences that were statistically significant at the 95 per cent level and ignored the estimated differences for the 'N/A' categories. All of the shortlisted factors showed statistically significant estimated differences for all categories.

Table 16: Maximum estimated differences for the shortlisted factors (Full-time)

Factor	All factors	Option 1 (ABCS but no other student characteristics)	Option 2 (student characteristics but no ABCS)
ABCS progression quintile	4.1	6.1	N/A
Age on entry	4.3	N/A	6.3
Deprivation quintile (IMD)	1.8	N/A	3.4
Disability	1.9	N/A	2.4
Entry qualifications	12.0	11.8	12.5
Ethnicity	3.2	N/A	3.9
Geography of earnings quintile	23.3	22.8	23.2
Level of study	9.2	8.0	9.7
Sex	2.3	N/A	3.0
Subject of study (CAH level 2)	27.4	27.0	28.2
Year	1.1	1.1	1.2

Factor	All factors	Option 1 (ABCS but no other student characteristics)	Option 2 (student characteristics but no ABCS)
Total possible distinct benchmarking groups	26,438,400	81,600	6,609,600

69. From the data in Table 16 we can see that **subject of study**, **geography of earnings quintile**, **entry qualifications** and **level of study** are factors with the largest estimated differences and the strongest explanatory power, and we therefore consider we should prioritise these factors. Including the geography of earnings quintile ensures that the benchmark takes some account of how the outcome is affected by the area where the graduate lives and works.

70. **Qualifying year** was prioritised in the progression measure in the past and was considered a factor because some years of data may have been impacted by the pandemic. LEO data for both 2020-21 and 2021-22 is used for the 2017-18 and 2018-19 qualifying years, and LEO data for 2020-21 is used for the 2016-17 qualifying year. We can see that qualifying year has limited explanatory power for the earnings measure, but we propose to continue to include it for consistency with the other post-study measures, to ensure that any significant changes to the labour market are accounted for and to support comparability for providers assessed in different years.

71. The results also show that student characteristics of age on entry, deprivation quintile (IMD), disability, ethnicity and sex are significant in Option 2 and ABCS is significant in Option 1. As the separate characteristics are all reflected in the ABCS measure, we consider that for consistency with progression and use of skills, Option 1 is preferable.

72. To understand how implementing Option 1 would impact the benchmarking process we have looked at the number of benchmarking groups, percentage of students in very small groups and the contribution of providers to their own benchmarks.

73. Table 17 shows the number of categories for each of the factors that were included in the modelling for the full-time mode of study for Option 1 and Table 18 shows this for Option 2. Table 19 shows the extent of each provider's contribution to its own benchmark to give an indication of the risk of self-benchmarking.

Table 17: Categories of benchmarking factors (Option 1)

Factor	Count
ABCS progression quintile	3
Entry qualifications	10
Geography of earnings quintile	5
Level of study	3
Subject of study (CAH level 2)	34
Year	4

Factor	Count
Summary	
Number of possible combinations	81,600
% of students in very small groups (1 to 5)	3%
% of students in small groups (1 to 20)	13%

Table 18: Categories of benchmarking factors (Option 2)

Factor	Count
Age on entry	3
Deprivation quintile (IMD)	2
Disability	2
Entry qualifications	10
Geography of earnings quintile	5
Ethnicity	4
Level of study	3
Sex	2
Subject of study (CAH level 2)	34
Year	4
Summary	
Number of possible combinations	6,609,600
% of students in very small groups (1 to 5)	28%
% of students in small groups (1 to 20)	53%

Table 19: Providers' contributions to their own benchmarks

Percentage of providers with at least (X%) contribution to their own benchmark	Earnings (Full-time)		
	>5%	>20%	>50%
Option 1 (ABCS but no other student characteristics)	90%	4%	0%
Option 2 (Student characteristics but no ABCS)	100%	86%	4%

74. For the set of factors proposed in Option 1, our view is that the number of benchmarking groups is manageable, minimises the number of students in very small groups, and does not

result in too much self-benchmarking. We therefore consider that the judgements made about the prioritisation of the shortlisted factors are appropriate given the modelling results and the inclusion of the intersectional ABCS factor. We consider that Option 3 achieves the most appropriate balance of policy objectives and statistical properties of the three options and propose it is therefore consistent with the principles for the selection and application of benchmarking factors.

Analysis of benchmark factors for part-time students

75. For part-time students, we started by carrying out the same steps as for full-time students, using with the same factors as used for progression. However, because there are more students included in the LEO data than there are in the Graduate Outcomes survey data, we are able to use more detailed groupings within some factors. The analysis in this section uses more detailed groupings for both subject and the geography of earnings. These groupings are explained in paragraphs x-y.

76. Table 20 shows the largest significant estimated difference for each of the shortlisted factors. It aims to provide an indication of the relative correlation with the outcome of the earnings measure. When compiling the table, we only compared estimated differences that were statistically significant at the 95 per cent level and ignored the estimated differences for the 'N/A' categories. All of the shortlisted factors except age on entry showed statistically significant estimated differences for all categories, including the 'N/A' categories. The '21-30' category for age on entry was the only category with an estimated difference that was not statistically significant. Note that we considered the differences across all of the categories, as well as the numbers of students associated with each category, when making our judgements.

Table 20: Maximum estimated differences for the shortlisted factors (Part-time)

Factor	All factors	Option 1 (ABCS but no other student characteristics)	Option 2 (student characteristics but no ABCS)
ABCS progression quintile	0.9	8.2	N/A
Age on entry	3.5	N/A	3.8
Deprivation quintile (IMD)	1.0	N/A	1.2
Disability	3.5	N/A	3.5
Entry qualifications	5.1	5.4	5.1
Ethnicity	6.3	N/A	6.4
Geography of earnings quintile	21.2	20.9	21.3
Level of study	4.8	4.1	4.8
Sex	10.9	N/A	11.1
Subject of study (CAH level 2)	39.0	43.4	39.0
Year	1.3	1.5	1.6

77. From the data in Table 20 we can see that **subject of study** and **geography of earnings quintile** are factors with the largest estimated differences and the strongest explanatory power and are factors that should be prioritised.

78. While **entry qualifications** and **level of study** have lower explanatory power than some other factors, we consider that these factors should be prioritised for consistency with the factors selected for full-time students, if this does not impact the risk of self-benchmarking. We also propose to include **qualifying year** for consistency with full-time and for the same reasons.

79. Table 20 suggests that the **ABCS progression quintile** captures much of the variation explained by the individual characteristics and has high explanatory power once the other personal characteristics are removed. But there could be an argument that Option 2 captures more variation than Option 1. When all factors are included, the ABCS factor has a smaller estimated maximum difference than some of the other personal characteristics it's replacing. This is different from the full-time earnings measure and all other measures where we are proposing to replace personal characteristics with an ABCS factor: usually the ABCS factor is clearly bringing more explanatory power than the other factors. For this reason, we consider both Option 1 and Option 2 in step 2 to assess whether one option is preferable to the other.

80. Table 21 shows the numbers of categories for each of the factors that were included in the modelling for the part-time mode of study. All the factors except entry qualifications and sex have the same categories as those used for modelling the full-time mode of study.

Table 21: Categories of the shortlisted factors (Part-time)

Factor	Count
Both options	
Entry qualifications	3
Geography of earnings quintile	5
Level of study	3
Subject of study (CAH level 2)	34
Year	4
Option 1 only (ABCS progression instead of student characteristics)	
ABCS progression quintile	4
Option 2 only (student characteristics instead of ABCS quintile)	
Age on entry	3
Deprivation quintile (IMD)	2
Disability	2
Ethnicity	5
Sex	2

81. Table 22 shows the maximum possible number of benchmark groups for each option, Table 23 shows the percentage of students in each size of benchmarking factor group, and Table 24 shows the extent of each provider's contribution to their own benchmark. These tables can help us to determine whether Option 1 or Option 2 is preferable.

Table 22: Maximum possible number of benchmark factor groups for each option (Part-time)

Option 1 (ABCS progression instead of student characteristics)	Option 2 (Student characteristics instead of ABCS progression)
24,480	1,321,920

Table 23: Percentage of students in each size of benchmark factor group (Part-time, Apprenticeship)

Mode	Benchmark factor group size	Option 1 (ABCS progression instead of student characteristics)	Option 2 (Student characteristics instead of ABCS progression)
Part-time	1 to 5 students	7%	37%
Part-time	1 to 20 students	24%	62%
Apprenticeship	1 to 5 students	11%	35%
Apprenticeship	1 to 20 students	28%	65%

Table 24: Percentage of providers in each contribution bucket (Part-time, Apprenticeship)

Mode	Contribution to benchmark	Option 1 (ABCS progression instead of student characteristics)	Option 2 (Student characteristics instead of ABCS progression)
Part-time	≤ 5%	20%	0%
Part-time	> 5% and ≤ 20%	57%	34%
Part-time	> 20% and ≤ 50%	19%	46%
Part-time	> 50%	4%	20%
Apprenticeship	≤ 5%	14%	0%
Apprenticeship	> 5% and ≤ 20%	49%	19%
Apprenticeship	> 20% and ≤ 50%	24%	54%
Apprenticeship	> 50%	14%	27%

82. From these tables we can see that Option 1 results in many fewer benchmarking groups. With Option 2 there would be:

a. more students in the smallest benchmarking groups (for part-time students, 37 per cent compared with 7 per cent for Option 1), and

b. more providers contributing to at least 20 per cent of their own benchmark (for part-time students, 66 per cent of providers compared with 23 per cent for Option 1).

83. Option 1 also has the advantage of consistency with other measures, and incorporating a wider range of factors, including intersection between characteristics, than Option 2. We therefore propose to use the factors included in Option 1.

Grouping subject and geography factors for the part-time and apprenticeship earnings measure

84. There are more graduates included in the LEO data than there are in the Graduate Outcomes survey data. This is largely because LEO data does not rely on responses to a survey.

85. More students in the population for the earnings measures means that we have the chance to use more detailed categories within the benchmarking factors to improve the accuracy of the benchmarks, without increasing the level of self-benchmarking to an unsustainable level. The two factors we consider most useful to show in more detail for the earnings measure are subject group and the geography of earnings quintile, because these factors have the greatest explanatory power and are used in more detail in the full-time measure.

86. Looking across these two factors, there are four options for the approach we could take:

- **Option 1: detailed subject groups.** This is the option we are proposing. The progression measure uses a broad subject grouping as the benchmarking factor for part-time and apprenticeship students, with 10 categories. Using the more detailed CAH level 2 grouping for the earnings measure for part-time and apprenticeship students would instead mean there were 34 categories. This would mean graduates would be more often compared with other graduates doing a more similar course to them. For instance, one broad subject group is ‘Law and social sciences’. We could instead use more detailed subject groups, including ‘Law’, ’Economics’ and ‘Health and social care’.

- **Option 2: detailed geography groups.** The progression measure combines some of the geography of employment quintiles into three groups rather than five. Using all five quintiles as separate groups for benchmarking the earnings measure would mean that graduates were more often compared against graduates from similar types of area. For instance, quintile 1 is combined with quintile 2. We could instead separate out these quintiles.

- **Option 3: broad groups for both.** This would use factors with the same granularity as the progression and use of skills measures for part-time and apprenticeship students.

- **Option 4: detailed subject groups and detailed geography groups.**

87. We do not consider it appropriate to use detailed groups for both (Option 4) because it would create too many distinct benchmarking groups and increase self-benchmarking to a level where too many providers are contributing most of their own benchmarks.

88. Our view is that detailed subject groups will have a more positive impact than using detailed geography of earnings groups. We recognise that, when looking at earnings of graduates, some of the subjects within the broad group can have different earnings expectations. For instance, nursing and psychology are in the same broad subject group, and our modelling shows that there is a marked difference in the proportion earning above the threshold we propose to use, even after taking other factors into account. Using the detailed subject grouping would therefore reduce the risk of a provider's students being compared with those whose courses are not as similar. While this is also true to some extent for grouping geography quintiles together, the differences between quintiles in the same group is not as great, and it is less likely that a whole provider would be impacted by geography groups being combined.

89. We therefore proceeded to consider which geography of earnings quintiles to combine and the impact this has on the size of the benchmarking groups.

Combining geography of earnings quintiles

90. We calculated the earnings measure for each geography of earnings quintile and mode of study. The proportion of graduates counted as a positive outcome in each quintile is shown in Table 25.

Table 25: Earnings measure by geography of earnings quintile

Geography of earnings quintile	All modes	Full-time	Part-time	Apprenticeship
1 (lowest)	52%	49%	70%	71%
2	60%	57%	75%	74%
3	64%	62%	76%	75%
4	71%	70%	78%	86%
5 (highest)	75%	75%	81%	84%

91. It is clear that quintile 1 should not be grouped with any other quintiles due to the large difference between the proportions for quintile 1 and the proportions for quintile 2 (8 percentage points across all modes). For part-time and apprenticeship students, the differences between quintile 2 and quintile 3 are smaller than the differences between quintile 1 and quintile 2, and also smaller than the differences between quintile 3 and quintile 4. Therefore we propose to combine quintile 2 with quintile 3, and quintile 4 with quintile 5.

92. Our proposed groupings for part-time and apprenticeship students are therefore:

- Quintile 1
- Quintiles 2 and 3
- Quintiles 4 and 5.

Impact on benchmarking group sizes

93. Tables 26 and 27 show the effect of grouping these quintiles on the size of the benchmarking factor groups and the contributions to the benchmarks made by individual providers for the overall indicators.

Table 26: Effect of grouping the geography of earnings quintiles on the percentage of students in each size of benchmark factor group

Mode	Benchmark factor group size	Quintiles as five separate groups	Quintiles in three groups
Part-time	1 to 5 students	7%	4%
Part-time	1 to 20 students	23%	15%
Apprenticeship	1 to 5 students	11%	7%
Apprenticeship	1 to 20 students	28%	22%

Table 27: Effect of grouping the geography of earnings quintiles on the percentage of providers in each contribution category

Mode	Contribution to benchmark	Quintiles as five separate groups	Quintiles in three groups
Part-time	≤ 5%	20%	36%
Part-time	> 5% and ≤ 20%	57%	49%
Part-time	> 20% and ≤ 50%	19%	11%
Part-time	> 50%	4%	3%
Apprenticeship	≤ 5%	14%	24%
Apprenticeship	> 5% and ≤ 20%	49%	47%
Apprenticeship	> 20% and ≤ 50%	24%	20%
Apprenticeship	> 50%	14%	9%

94. Table 26 shows that reducing from five quintile groups to three means that the proportion of students in very small benchmarking groups is reduced (to 4 per cent for part-time, 7 per cent for apprenticeship). Table 27 shows that reducing to three groups means a smaller proportion of providers have high contributions to benchmark: the proportion with a contribution of over 20 per cent drops from 23 per cent to 14 per cent for part-time, and 38 per cent to 29 per cent for apprenticeship.

95. At the overall measure level, we could tolerate using five separate quintiles, but given the importance of split indicators and meaningful benchmarks at that level, we are proposing to reduce to three groups. When looking at split indicators, all the benchmarking groups are divided up into smaller parts (for instance, when creating a benchmark for students aged under 21, all benchmarking groups will only include students aged under 21). This will

increase the proportion of students in small benchmarking groups, and the proportion of providers with high levels of self-benchmarking. For this reason, we consider that using three groups for the geography of earnings quintiles for part-time and apprenticeship students will achieve the best balance in terms of the policy objectives of benchmarking and the need to maintain statistical robustness.

ABCS measures for NSS

96. This section sets out the approach we have taken to develop two new Association between characteristics of students (ABCS) measures – one for full-time students and one for part-time students – to support the benchmarking of the NSS.

97. ABCS is a set of measures that seeks a better understanding of how outcomes vary for groups of students holding different sets of characteristics. We define groups of students by looking at a set of characteristics so that we can determine the relationship of not just one characteristic on an outcome, but the relationship of multiple characteristics.

98. Each measure takes the form of a set of quintiles, where quintile 1 indicates groups of students who are least likely to experience or report a positive outcome and quintile 5 indicates group that are most likely to experience or report a positive outcome.

99. These quintiles are defined using the outcomes from statistical modelling. Details of the modelling methodology and creation of quintiles can be found in the associated methodology document.[10]

100. ABCS full-time (FT) and part-time (PT) NSS measure the proportion of students responding positively to the National Student Survey (NSS). This is based on their response to the question ‘How well have teaching staff support your learning?’. This question has been chosen as levels of positivity for this question have been shown to be highly correlated with other questions across the survey, and more consistently than any other question. Students responding ‘Very well’ or ‘Well’ are considered to have a positive response.

101. The measures are separate for full-time and part-time students because the demographics of full-time and part-time students are quite different and so the relationship between various characteristics and the chances of giving positive responses differs between these modes of study.

Population

102. This analysis was carried out before the publication of the NSS 2026 results and used the three years of available data to construct the measures. Updates were made to the NSS for students taking the survey in 2023, and results were not comparable with earlier versions of the survey. Therefore, the model used to create the ABCS FT and PT NSS included those taking the survey in 2023, 2024 and 2025.

103. The measure is further restricted to UK domiciled students studying at English providers who responded to Question 16 of the NSS. Students domiciled outside the UK before their course are excluded because they are out of scope for many of the characteristics included in the

[10] See Associations between characteristics of students - Office for Students.

ABCS measures. Students who are not UK-domiciled would therefore be a separate benchmarking group.

Selection of characteristics

104. We started with the shortlist of factors selected for the NSS benchmarking considered previously for the NSS publication and for use in TEF.[11]

105. The characteristics used for ABCS FT NSS are age group, disability type, eligibility for free school meals, estrangement, ethnicity, sex, sexual orientation and religious belief.

106. The characteristics used for ABCS PT NSS are age group, disability type, ethnicity, parental education, income deprivation affecting children index (IDACI), sex, and local or distance learners.

The statistical model

107. We have used a binary logistic regression model to quantify the relationship between student characteristics and observed positive NSS responses for both full-time and part-time students. This relationship is referred to as a 'predicted rate' for an individual given their characteristics, that is the likelihood of someone with those characteristics having responded positively to the NSS in our historical data. We have included all characteristics as main effects and used a statistical approach to determine which of the two-way interactions should be included.[12] See the ABCS methodology document for details.[13]

108. Using the model, we calculate the predicted positive response rates for each group of respondents. A group is defined as a set of individuals who have all of their characteristics in common. For example, if we only include age and sex in the model, all 19 year-old females would form one 'group'.

Derivation of ABCS quintiles

109. We then used these predicted rates to split the individuals included in the modelling into five quintiles. Those groups with the lowest predicted rates are in the lowest quintiles and those with the highest are in the highest quintile. Table 28 shows the number and proportion of graduates in each quintile for the full-time measure, as well as the mean, minimum and maximum predicted rate. Table 29 does the equivalent for the part-time model. The minimum predicted rates are also the breakpoints, which determine the quintile boundaries.

[11] For more information on which characteristics are considered as candidates for benchmarking factors, see Review of the selection and grouping of benchmarking factors - Office for Students. Personal characteristics that made the shortlist of possible benchmarking factors was the starting point for assessing characteristics for inclusion in the ABCS measure.

[12] More details on the characteristics included and the two-ways interactions can be found in the ABCS NSS documentation, available at: Associations between characteristics of students - Office for Students.

[13] See Associations between characteristics of students - Office for Students.

Table 28: Description of ABCS FT NSS quintiles

FT NSS quintile	Number of students	Proportion of students (%)	Mean modelled FT positivity rate (%)	Minimum modelled FT positivity rate (%)	Maximum modelled FT positivity rate (%)
Quintile 1	134,082	19.99	81.3344	53.3420	83.1386
Quintile 2	134,227	20.01	84.0369	83.1392	84.9905
Quintile 3	134,187	20.00	85.4294	84.9925	85.9462
Quintile 4	134,108	19.99	86.6020	85.9465	87.1677
Quintile 5	134,228	20.01	89.5898	87.1679	98.2904

Table 29: Description of ABCS PT NSS quintiles

PT NSS quintile	Number of students	Proportion of students (%)	Mean modelled PT positivity rate (%)	Minimum modelled PT positivity rate (%)	Maximum modelled PT positivity rate (%)
Quintile 1	10,202	19.54	80.3870	55.6198	82.7773
Quintile 2	10,678	20.45	83.8455	82.8060	84.6918
Quintile 3	10,414	19.95	85.5493	84.6958	86.3994
Quintile 4	10,458	20.03	87.2111	86.4007	88.0294
Quintile 5	10,455	20.03	89.6343	88.0357	99.1663

Defining a material difference from benchmarks

110. To support consistent interpretation of the TEF data indicators in relation to benchmarks, we have previously defined materiality values that allow assessors to interpret whether indicator performance is:

- materially above benchmark
- broadly in line with benchmark
- materially below benchmark.

111. We set a materiality value of 2.5 percentage points for all indicators for TEF 2023. This means that performance 2.5 per cent or more above benchmark (with enough statistical confidence) would be judged as materially above benchmark. Performance 2.5 per cent or more below benchmark (with enough statistical confidence) would be judged as materially below benchmark. The analysis we carried out to inform setting the value at 2.5 per cent and the relevant considerations are available on the OfS website.[14] In summary, in setting this materiality value we sought to achieve a distribution that would allow assessors to differentiate between levels of performance while recognising that it would be unreasonable to consider very small differences from a provider's benchmark as being material enough to affect its assessment. We have also sought to avoid unnecessary complexity by using the same materiality value for all indicators.

112. For this consultation, we have undertaken new analysis to understand if 2.5 per cent remains an appropriate materiality value, particularly in light of the introduction of additional post-study measures. This involved comparing the current 2.5 percentage point threshold against alternatives of 3 and 5 percentage points.

113. We conducted analysis for both student outcomes and NSS indicators across all modes of study, with additional breakdowns by provider size. We placed providers into one of five categories, reflecting where at least 80 per cent of the uncertainty distribution of the indicator is either:

a. Materiality below benchmark.

b. Crossing materially below and broadly in line with benchmark.

c. Broadly in line with benchmark.

d. Crossing materially above and broadly in line with benchmark.

e. Materially above benchmark

114. We placed providers into a fifth category (f) if 80 per cent of the uncertainty distribution of the indicator spans all three areas. It would not be possible to place weight on indicators in category f when reaching judgements.

[14] See Materiality and high benchmark values for use in interpretation of the TEF indicators - Office for Students.

115. Tables 30-33 below show the percentage of providers in each category for the three values we tested. They show the results for all providers for full-time and part-time modes of study.

116. The analysis showed that the existing 2.5 percentage point materiality value produces a reasonable distribution of providers that are materially below, broadly in line with, or materially above their benchmarks. Additionally, it does not result in large numbers of providers crossing three areas. We consider this distribution is reasonable across all indicators, including the new post-study indicators, and we found this to be the case when we limited the analysis to larger providers.

117. We considered increasing the values to 3 or 5 percentage points; however, this would narrow the range of outcomes treated as materially different from benchmark. This would reduce the ability of the indicators to help compare and differentiate performance across providers. We do not view a reduction in differentiation as desirable given the aims of the TEF to drive improvement and recognise and reward excellence. We therefore propose to retain the current 2.5 percentage point materiality value and to apply it across all indicators.

Table 30: Full-time – distribution of providers for student outcomes measures (all providers)

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Completion (proposed option)	**364**	**2.5**	**19.5**	**12.9**	**29.7**	**13.7**	**21.7**	**2.5**
Completion	364	3	17.9	10.4	39.0	13.2	17.9	1.6
Completion	364	5	10.7	8.8	60.4	9.9	10.2	0.0
Continuation (proposed option)	**372**	**2.5**	**14.0**	**13.4**	**33.3**	**18.3**	**20.7**	**0.3**
Continuation	372	3	12.1	10.5	43.3	16.7	17.5	0.0
Continuation	372	5	8.1	5.4	65.9	11.8	8.9	0.0
Progression (proposed option)	**339**	**2.5**	**20.6**	**16.2**	**21.8**	**13.9**	**20.4**	**7.1**
Progression	339	3	17.4	15.9	26.0	16.5	16.5	7.7
Progression	339	5	10.3	15.9	43.4	17.4	9.7	3.2
Use of skills (proposed option)	**334**	**2.5**	**9.0**	**17.7**	**23.7**	**21.3**	**18.9**	**9.6**
Use of skills	334	3	6.9	14.7	31.1	21.6	17.7	8.1
Use of skills	334	5	1.8	14.1	50.3	21.3	9.9	2.7
Earnings (proposed option)	**276**	**2.5**	**28.6**	**16.3**	**20.7**	**15.9**	**12.7**	**5.8**
Earnings	276	3	25.4	17.8	28.3	13.4	11.2	4.0

Earnings	276	5	14.9	17.0	48.6	11.6	6.2	1.8

Table 31: Part-time – distribution of providers for student outcomes measures (all providers)

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Completion (proposed option)	**254**	**2.5**	**18.5**	**11.0**	**4.3**	**20.1**	**39.0**	**7.1**
Completion	254	3	17.3	10.6	7.9	20.9	36.6	6.7
Completion	254	5	13.8	8.3	26.4	20.1	27.6	3.9
Continuation (proposed option)	**256**	**2.5**	**16.4**	**12.9**	**3.5**	**16.4**	**41.8**	**9.0**
Continuation	256	3	15.6	13.7	5.9	19.1	38.7	7.0
Continuation	256	5	11.7	12.5	23.0	25.0	27.0	0.8
Progression (proposed option)	**186**	**2.5**	**17.2**	**15.1**	**3.8**	**29.0**	**28.5**	**6.5**
Progression	186	3	15.1	17.2	6.5	32.3	24.2	4.8
Progression	186	5	10.8	17.2	17.7	38.2	11.3	4.8
Use of Skills (proposed option)	**179**	**2.5**	**7.8**	**19.0**	**1.7**	**30.7**	**20.7**	**20.1**
Use of Skills	179	3	6.7	22.9	3.4	36.9	16.8	13.4
Use of Skills	179	5	5.6	22.9	21.2	35.8	8.4	6.1

Earnings (proposed option)	**219**	**2.5**	**22.4**	**25.1**	**9.6**	**21.0**	**11.9**	**10.0**
Earnings	219	3	21.5	26.9	16.0	20.1	10.0	5.5
Earnings	219	5	11.4	26.5	37.4	16.4	5.0	3.2

Table 32: Full-time – distribution of providers for student experience measures (all providers)

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Academic support (proposed option)	**331**	**2.5**	**14.5**	**12.7**	**31.4**	**24.2**	**13.3**	**3.9**
Academic support	331	3	12.4	11.5	42.3	23	8.5	2.4
Academic support	331	5	6	12.1	65.6	12.7	2.4	1.2
Assessment and feedback (proposed option)	331	2.5	16.9	14.2	20.2	24.5	19.3	4.8
Assessment and feedback	331	3	13.6	14.2	28.1	24.5	16	3.6
Assessment and feedback	331	5	8.2	11.5	55	18.1	5.1	2.1
Learning opportunities (proposed option)	**331**	**2.5**	**10.6**	**14.5**	**32.9**	**20.8**	**15.1**	**6**

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Learning opportunities	331	3	8.2	13.9	40.8	19	12.4	5.7
Learning opportunities	331	5	3.6	10	64	15.4	5.1	1.8
Learning resources (proposed option)	331	2.5	28.1	15.4	27.2	15.1	10.9	3.3
Learning resources	331	3	26.3	16	32.9	13.9	8.2	2.7
Learning resources	331	5	18.4	13.6	56.5	6.6	3.3	1.5
Teaching on my course (proposed option)	**331**	**2.5**	**10**	**16**	**30.5**	**24.5**	**13.3**	**5.7**
Teaching on my course	331	3	8.2	14.5	42.3	22.1	10	3
Teaching on my course	331	5	4.8	7.6	66.2	16.6	2.4	2.4
Student voice (proposed option)	331	2.5	22.4	15.1	22.7	16.9	17.5	5.4
Student voice	331	3	21.1	14.5	28.7	16.9	14.2	4.5
Student voice	331	5	17.2	11.5	48.9	13.6	7.3	1.5

Table 33: Part-time – distribution of providers for student experience measures (all providers)

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Academic support (proposed option)	**90**	**2.5**	**26.7**	**17.8**	**5.6**	**27.8**	**15.6**	**6.7**
Academic support	**90**	3	22.2	23.3	8.9	27.8	13.3	4.4
Academic support	**90**	5	14.4	18.9	31.1	27.8	5.6	2.2
Assessment and feedback (proposed option)	**90**	**2.5**	27.8	22.2	8.9	28.9	4.4	7.8
Assessment and feedback	**90**	3	26.7	22.2	12.2	27.8	2.2	8.9
Assessment and feedback	**90**	5	20	23.3	37.8	15.6	2.2	1.1
Learning opportunities (proposed option)	**90**	**2.5**	**19.8**	**30.8**	**4.4**	**29.7**	**6.6**	**8.8**
Learning opportunities	**90**	3	19.8	28.6	7.7	30.8	5.5	7.7
Learning opportunities	**90**	5	12.1	20.9	36.3	24.2	2.2	4.4
Learning resources (proposed option)	**90**	**2.5**	36.7	28.9	3.3	13.3	6.7	11.1

Measure	Total providers	Materiality value (%)	a. Materially below (%)	b. Crossing materially below and broadly in line with benchmark (%)	c. Broadly in line with benchmark (%)	d. Crossing materially above and broadly in line with benchmark (%)	e. Materially above (%)	f. Crossing all 3 areas (%)
Learning resources	**108**	3	31.1	31.1	10	13.3	4.4	10
Learning resources	**108**	5	22.2	26.7	36.7	12.2	2.2	0
Teaching on my course (proposed option)	**108**	**2.5**	**25.6**	**20**	**6.7**	**33.3**	**5.6**	**8.9**
Teaching on my course	**108**	3	24.4	20	10	31.1	5.6	8.9
Teaching on my course	**108**	5	16.7	17.8	47.8	13.3	2.2	2.2
Student voice (proposed option)	**108**	**2.5**	12.2	14.4	2.2	22.2	33.3	15.6
Student voice	**108**	3	12.2	16.7	3.3	26.7	30	11.1
Student voice	**108**	5	8.9	14.4	20	26.7	24.4	5.6

Very high absolute thresholds

118. In the main consultation document (Proposal 13), we propose that very high absolute continuation or completion rates of at least 95 per cent would indicate outstanding performance, even where benchmarks have not been exceeded. Previously in the TEF, we used a benchmark-based rule, where if a provider had a benchmark value above 95 per cent, we treated performance in line with benchmark as outstanding.

119. To inform this proposal, we tested very high absolute thresholds of 80, 85, 90, 95 and 97 per cent and compared the number of indicators that would be 'reclassified' as outstanding using each of these to the results of applying the previous rule. Indicators were reclassified if 90 per cent of the uncertainty distribution around the indicator was above the relevant threshold. We analysed distribution charts and reclassification tables across all student outcome measures and modes of study to understand patterns of absolute indicator values and the impact of applying each threshold, additionally considering any differential impact for smaller providers.

120. Figures 1-4 below show the distribution for the top quartile of all providers' indicator values for each of continuation and completion measures, for the full-time and part-time mode of study. Circles represent providers who are broadly in line with their benchmark, and triangles those that are materially above their benchmark. They allow us to see how many providers have high absolute indicator values, but are not exceeding their benchmark. Table 34 then shows how many of these would be considered as outstanding if we used 90, 95 and 97 per cent as a high absolute threshold, and how this compares with the number that would benefit from the previous benchmark-based rule.

121. As Table 34 shows, the lower values that we tested resulted in many more indicators being reclassified than the current approach and we therefore did not consider that they would represent an appropriate bar for outstanding performance. Using a value of 95 per cent or greater, however, was only slightly more generous than the previous benchmark-based rule, and the outcomes of the two were broadly comparable. Using 97 per cent resulted in very little reclassification. We therefore consider that using a threshold of 95 per cent, above which performance would be considered outstanding best meets our aims of ensuring that we are recognising very high absolute levels of performance alongside those exceeding their benchmark. In no cases did the indicators that were reclassified fall below the provider's benchmark, so we do not view any of the reclassification as being in tension with a provider's performance against its benchmark.

122. We are proposing only to use this very high threshold for the continuation and completion indicators. This approach is consistent with using minimum absolute thresholds for the continuation and completion indicators, for assessing whether outcomes meet the requirements of condition B3. In general, this means that continuation and completion rates would be interpreted through a combination of absolute and benchmarked performance.

123. We do not propose to set very high absolute thresholds for the post-study measures, for the reasons set out in Proposal 13 of the consultation.

Figure 1: Full-time continuation rates for the quarter of providers with highest rates

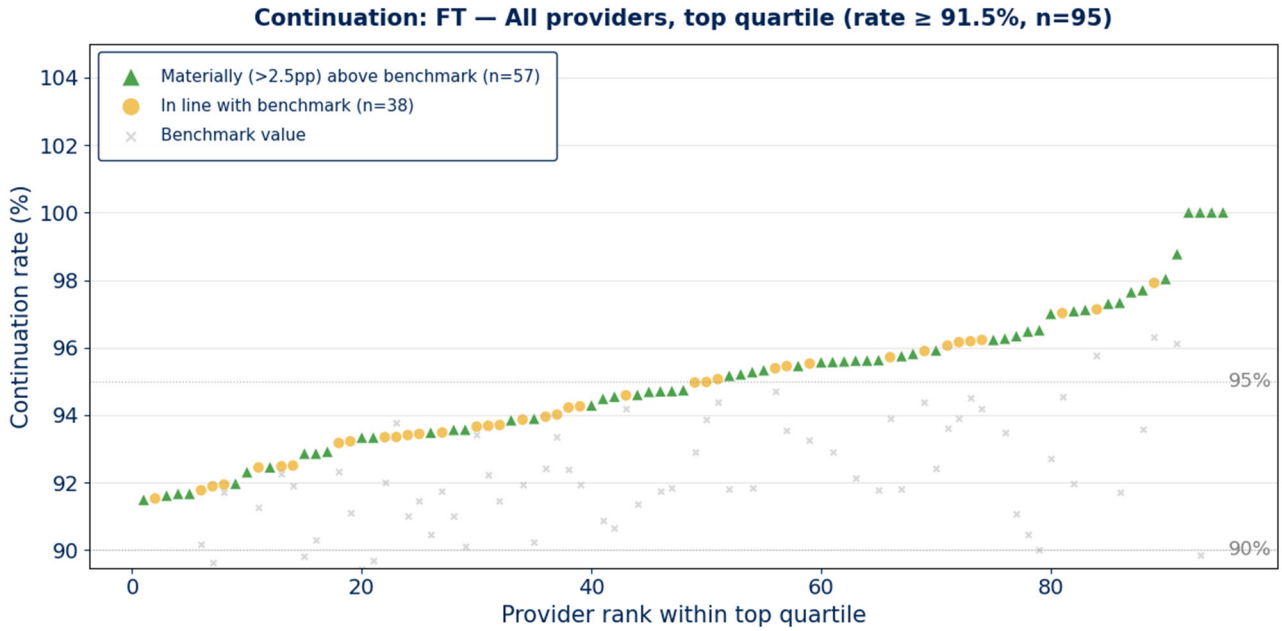


Figure 2: Part-time continuation rates for the quarter of providers with highest rates

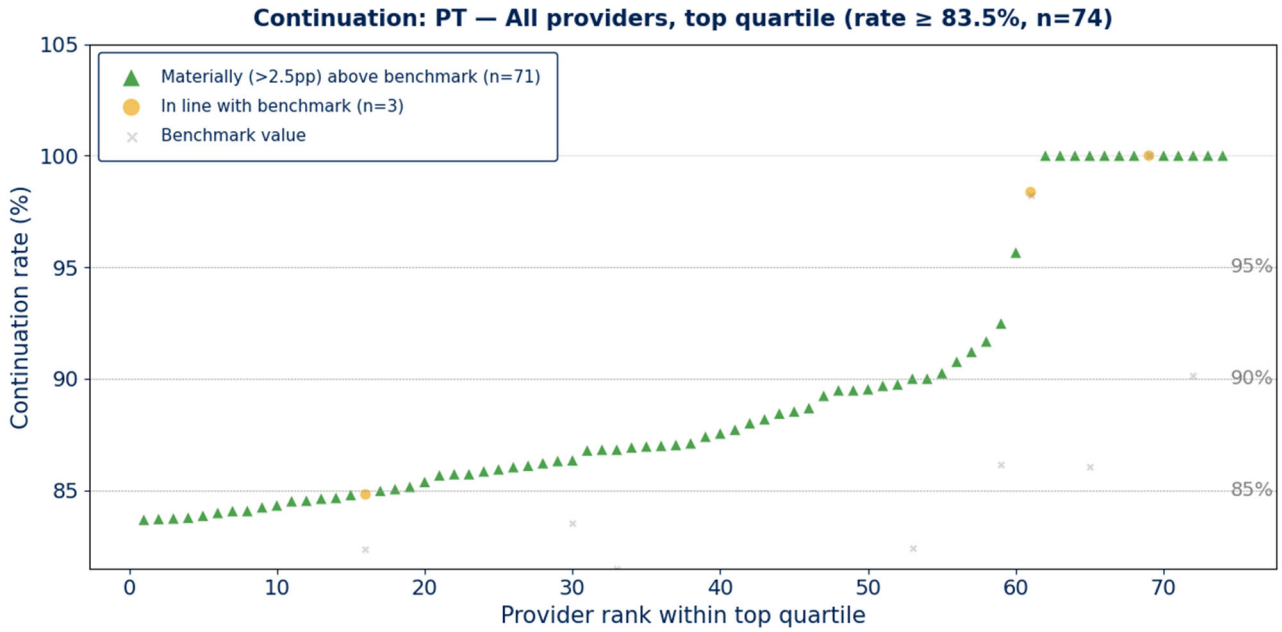


Figure 3: Full-time completion rates for the quarter of providers with highest rates

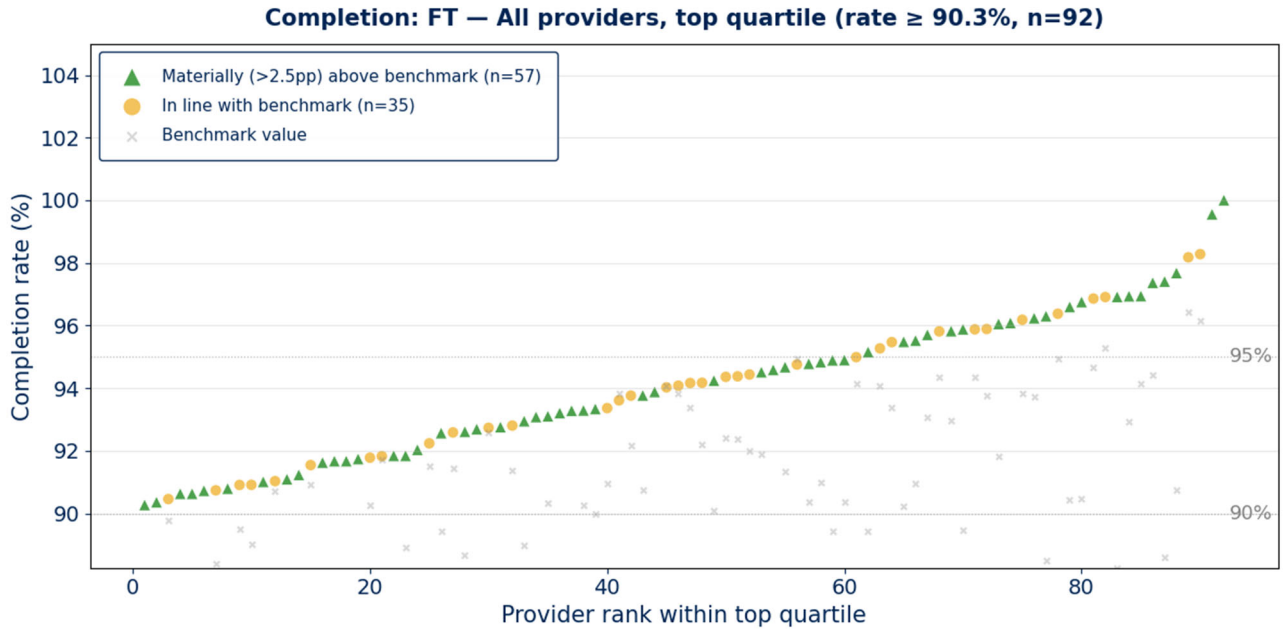


Figure 4: Part-time completion rates for the quarter of providers with highest rates

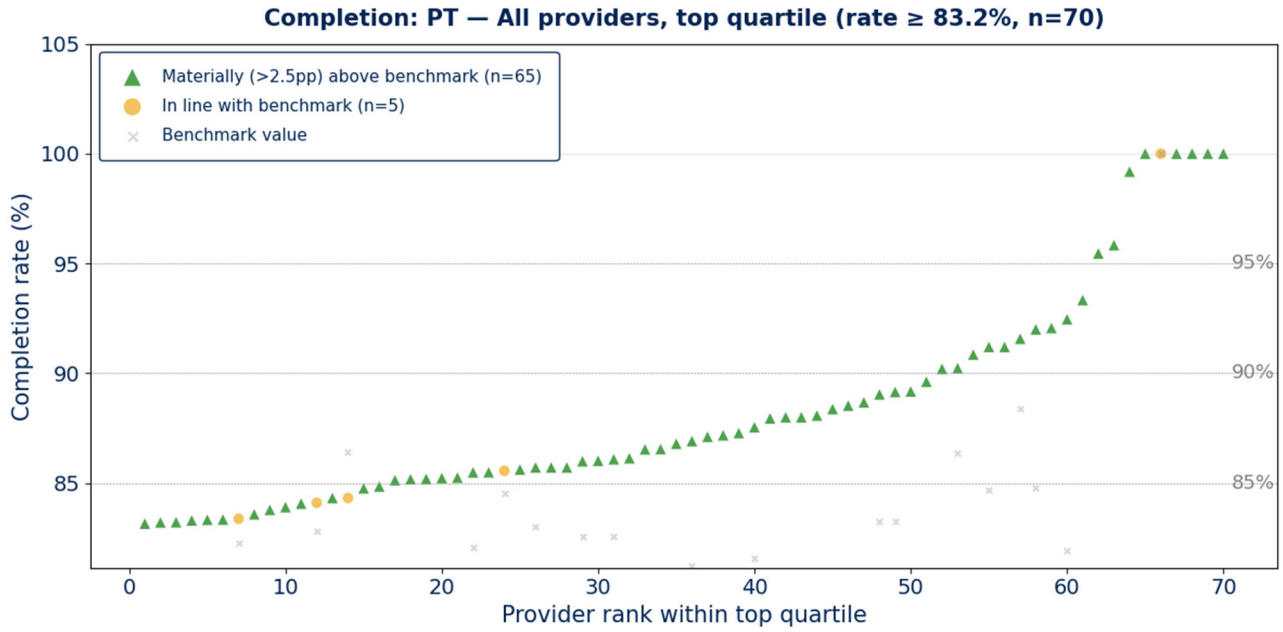


Table 34: Effect of very high absolute thresholds

Measure	Mode	Absolute threshold	Total providers	Number in line with benchmark treated as outstanding under new rule	Number in line with benchmark treated as outstanding under previous rule
Continuation	Full-time	90%	379	46	2
Continuation (proposed option)	**Full-time**	**95%**	**379**	**12**	**2**
Continuation	Full-time	97%	379	1	2
Continuation	Part-time	90%	295	1	2
Continuation (proposed option)	**Part-time**	**95%**	**295**	**1**	**2**
Continuation	Part-time	97%	295	1	2
Completion	Full-time	90%	366	34	3
Completion (proposed option)	**Full-time**	**95%**	**366**	**12**	**3**
Completion	Full-time	97%	366	2	3
Completion	Part-time	90%	277	1	1
Completion (proposed option)	**Part-time**	**95%**	**277**	**1**	**1**
Completion	Part-time	97%	277	1	1