National Student Survey 2023

Benchmarking in the NSS publication

Enquiries to nss@officeforstudents.org.uk

Publication date 10 August 2023

Contents

Introduction

1. The Office for Students (OfS) uses benchmarking to make meaningful comparisons between higher education providers, different student groups and other groups of interest. It is a method we use to take account of the mix of courses and students at a provider and indicate how well that provider has performed compared with performance for similar types of students on similar types of courses in the higher education sector as a whole.

2. Because there are differences between some of these groups and providers, average values for the whole of the higher education sector are not necessarily helpful when making these comparisons.

3. To account for some of the factors that contribute to these differences, the benchmark is calculated as a weighted sector 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 sector-overall rates for those student groups.

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, where we consider it would not be undesirable to control for those factors. These factors relate to characteristics of courses (such as subjects, and level of study) and students (such as their age and other personal characteristics).

5. The benchmark we intend can be used in two ways:

    - to understand how well a provider has performed compared with performance for similar types of students on similar types of courses in the higher education sector as a whole

    - to assess similarities between groups and providers.

6. **In practical terms**, the benchmarks included in the NSS publication can help users interpret the positivity measures calculated for each provider. If the positivity measure for a provider is much higher than the benchmark for a particular question, that means students at that provider were more positive about that aspect than ‘similar’ students across the sector. ‘Similar’ means students with the same subject, level and mode of study, and the matching personal characteristics across all the factors we use to create the adjusted sector average. Conversely, if the positivity measure for a provider is the same as the benchmark, students at that provider were equally positive about that aspect to similar students across the sector, and if the positivity measure is well below the benchmark, students were less positive than similar students across the sector.

7. This document outlines how we use benchmarking in the NSS publication, which factors are used for different groups or questions, what analysis and consideration led us to these decisions, and gives some of the technical detail for how the benchmarks are calculated.

8. In our approach to benchmarking in the NSS publication, we have drawn on research and consultation carried out to inform the student outcomes metrics and Teaching Excellence

Framework (TEF) indicators, also published by the OfS.[1] The approach to benchmarking described in this document mirrors the approach taken in these other publications.

[1] See, in particular: 'Description of student outcomes and experience measures used in OfS regulation', paragraphs 111 to 130; Annexes E, F, G and H, available at Description and definition of student outcome and experience measures - Office for Students. The approach used for the NSS aligns with the approach outlined, except for the focus on NSS questions and themes instead of indicators, the use of providers across all UK countries to define the sector (paragraph 14 in this document), and the change in approach for the CAH3 split (paragraphs 22-29).

Our benchmarking methodology

9. Our current methodology to benchmarking is well established and has been used in the UK Performance Indicators (UKPIs) since 1997.[2]

10. Technical details of the methodology are also described in 'Statistical analysis of performance indicators in UK higher education', Journal of the Royal Statistical Society.[3]

11. The methodology allows us to ask the question: 'What would the observed positivity rate have been at this provider if its distribution of students across benchmarking factor groups had been what it was, but its positivity measures across those same benchmarking groups were replaced by the sector-overall rates?'

12. When there are known differences between the experiences of some groups of students or providers, observed average values for the whole of the higher education sector are not necessarily helpful when forming this expectation. Instead, we calculate the benchmark as a weighted sector average, reflecting the number of students in that group at the provider. As such, benchmarks give information about the values that the sector overall might have achieved for the indicator if the characteristics included in the benchmarking factors are the only ones that are important. Where differences exist between an provider's positivity measure and its corresponding benchmark, these may be due to the provider's performance, or they may be due to some other characteristic which is not included in the weighting.

General approach to benchmarking

13. To create benchmarks for a given question or theme, we calculate the observed positivity measure for the higher education sector as a whole, for each benchmarking group. The benchmark for each provider is then calculated by taking a weighted average of the overall sector rates for each benchmarking group, taking account of the particular mix of students across those groups at the provider in question.

14. The benchmarking methodology used by the OfS means that a provider is not being compared with a pre-set group of providers, but rather the positivity measures for a provider's students are compared with the positivity measures of similar students across the entirety of the higher education sector. For the purpose of calculating these benchmarks for the NSS publication, the higher education sector within which we are making comparisons of the positivity measures for similar students is made up of all English higher education providers registered with the OfS, and all providers which are funded or regulated by one of the devolved administrations, at the time that we produce the indicators. This recognises the benefit of consistent UK-wide benchmarks to users of the NSS dashboard, and is a difference to the benchmarking approach used elsewhere by the OfS, where only English providers registered with the OfS are used for form the benchmarks for English providers.

[2] See UK Performance Indicators | HESA.

[3] D. Draper and others, 'Statistical analysis of performance indicators in UK higher education'. Journal of the Royal Statistical Society, Series A, volume 167, part 3, 2004.

For a worked example of how we calculate benchmarks, see Annex C: Worked example of benchmarking calculations.

For details of the formulae used in the calculation of benchmarks, see

Annex F: Technical detail about benchmarking calculations.

The benchmarking factors we use

15. The basis on which we select, define, and apply the factors used in benchmarking the NSS results is key to the integrity and robustness of the benchmark values calculated and interpreted by users.

16. The benchmarking factors we use is the result of us following the set of principles for the selection and application of benchmarking factors (see Annex E). In line with these principles, we have conducted a brief review of the benchmarking factors, as proposed in our consultation on the publication of results for the National Student Survey.[4] This analysis is set out in Annex A: Brief review of benchmarking factors – questions 1-27 and theme measures; and Annex B: Brief review of benchmarking factors – question 28, overall satisfaction. The outcome is that we are using the agreed factors for questions 1-27 and the theme measures, and using slightly fewer factors for benchmarking question 28.

17. We will conduct a fuller review of benchmarking factors to inform the benchmarking factors that will be used in the NSS 2024 and later publications.

18. In Tables 1 and 2 we summarise the benchmarking factors and groupings we are using for the NSS questions and theme scores, and compare them to the factors used for NSS 2022 and earlier. Note that these groupings are only for the purposes of benchmarking, to ensure that small groups of students can still contribute meaningfully to interpretation of a provider's performance, and do not imply that we consider the differences to be unimportant. For instance, students with 'other' sex will continue to be shown separately to female students where results are published by sex.

Table 1: New benchmarking factors (questions 1-27 and theme measures)

Factor	Previous factors	NSS 2023
Mode of study	Full-time, Part-time	Full-time, Part-time, Apprenticeship (Defined with reference to first year of engagement)
Subject of study	CAH level 1	For full-time students: CAH level 2 (with CAH19-02 Celtic studies combined with CAH19-04 Languages and area studies) For part-time students and apprenticeships: Broadly defined subject groups (see Annex D)
Level of study	Not used in benchmarking	Other undergraduate, First degree, Undergraduate with postgraduate components

[4] See Proposal 5: Benchmarks at Consultation on the approach to publication of results of the National Student Survey - Office for Students.

Factor	Previous factors	NSS 2023
Age on entry	Under 21, 21 to 24, 25 and over	Under 21 or unknown, 21 to 30, 31 and over
Ethnicity	Asian, Black, Other, White, Unknown	Asian, Black, Other, Unknown or White, Mixed, non-UK domiciled
Disability *(No change proposed)*	Declared, None known	Disability reported, No disability reported
Sex	Male, Female, Other	Female or Other, Male (only used for full-time students)

19. Table 2 shows the benchmarking factors used to create benchmarks for question 28 (overall satisfaction). This question is only asked to students at providers in Wales, Scotland and Northern Ireland, and the smaller number of responses means that we needed to reduce the number of benchmarking factors for the benchmarks to be meaningful and statistically robust. For more information on how this decision was reached, see Annex B: Brief review of benchmarking factors – question 28.

Table 2: New benchmarking factors (question 28)

Factor	Previous factors	NSS 2023
Mode of study	Full-time, Part-time	Full-time, Part-time, Apprenticeship (Defined with reference to first year of engagement)
Subject of study	CAH level 1	For full-time students: CAH level 2 (with CAH19-02 Celtic studies combined with CAH19-04 Languages and area studies) For part-time students and apprenticeships: Broadly defined subject groups (see Annex D)
Level of study	Not used in benchmarking	Other undergraduate, First degree, Undergraduate with postgraduate components
Age on entry	Under 21, 21 to 24, 25 and over	Under 21 or unknown, 21 to 30, 31 and over
Ethnicity	Asian, Black, Other, White, Unknown	Asian, Black, Other, Unknown or White, Mixed, non-UK domiciled (only used for full-time students)

Factor	Previous factors	NSS 2023
Disability *(No change proposed)*	Declared, None known	Disability reported, No disability reported (only used for full-time students)
Sex	Male, Female, Other	Female or Other, Male (only used for full-time students)

The 'contribution to own benchmark'

20. The contribution to own benchmark is the weighted average of the provider's own students contributing to the sector averages that are used to calculate their benchmark. It is calculated in percentage terms for each provider.

21. The contribution to own benchmark is a useful indicator of how many students the provider contributes to the benchmarking groups that their students are in. Higher contributions occur when the characteristics of students at the provider in question do not frequently occur among student populations in the wider sector, for example, when a provider is much larger than other providers with similar students, or when most of their students are in benchmarking groups with few other students. In either case, the greater the contribution to the benchmark, the more likely that the benchmark is close to the provider's positivity rate. This is because there is less other sector data that can provide the information necessary to make the benchmark a reliable estimate of the values that might have been expected for the provider. For providers with a high contribution to their own benchmark, their performance would need to be very different to the rest of the sector to get a sizeable difference from benchmark.

Modified approach for detailed subject (CAH level 3) benchmarks

22. We use three different levels of study when publishing the NSS results, ranging from the broadest (CAH level 1) to the most detailed (CAH level 3).[5]

23. Benchmarks are produced in the NSS data for the whole provider, but also for subgroups of the provider. We produce benchmarks for providers split by mode, level and subject of study.

24. Usually, when the OfS publishes a benchmark for a split, the whole sector used to generate the benchmark is limited to that split. For instance, if we were looking at a provider's apprenticeship students, only other apprenticeship students would be used as the comparator population.

25. However, for the most detailed subject group (CAH3), there would in many cases be too few students in the population to create meaningful benchmarks, and providers would often have high contributions to their own benchmarks. For this reason, when creating benchmarks for results split by CAH3, we have decided to restrict the comparator population using CAH2 subject groups, rather than using CAH3 subject groups to construct the sector averages.

[5] CAH stands for 'Common Aggregation Hierarchy', a standardised grouping of subject codes and terms. See Common Aggregation Hierarchy (CAH) | HESA.

26. For instance, if we are showing data for the CAH3 group ‘Statistics’, instead of only comparing with statistics courses across the UK, we will be comparing with all courses in the CAH2 subject area ‘Mathematical sciences’. So the statistics courses will also be compared to mathematics courses.

27. We have adjusted the ‘contribution to own benchmark’ for these splits, to be the greater out of:

    a. The contribution to own benchmark for the CAH2 group, and

    b. The contribution to own benchmark for the CAH3 group.

28. In most cases, (a) is greater, which means all students at the provider who contribute to the benchmark (that is, who are in the CAH2 group) are counted. In the example of the ‘Statistics’ CAH3 group, the contribution to own benchmark would be based on all students at the provider on either statistics or mathematics courses.

29. However, in some cases, (b) is greater. This means including students from the rest of the CAH2 group would make the contribution to own benchmark lower. This could happen when the students from the provider’s CAH3 group are quite different to others in the sector, so the contribution to own benchmark for that CAH3 group is very high. In such cases, including other students would make the contribution to own benchmark lower. This would be misleading, so we use (b) if that is greater.

Annex A: Brief review of benchmarking factors – questions 1 to 27 and theme measures

This annex is aimed at readers seeking more information about how we have decided on the benchmarking factors that are used to produce benchmarks for NSS results.

This annex describes the brief review carried out to determine which factors to use for all NSS questions and themes (except for question 28 which is covered in Annex B).

1. In the consultation on the approach to publication of results of the NSS[6] we proposed that we would conduct a brief review of the existing benchmarking factors. This review seeks to confirm that the statistical properties of the existing selection and grouping of benchmarking factors are not fundamentally different than those which have previously been observed in results from the 2022 and earlier NSS surveys. For the initial NSS 2023 publication, we would expect to make only minor modifications to the factors set out in Table 2 of the consultation document (such as an adjustment to the groupings used), or no changes.

2. For this review, our evaluation of the benchmarking factors has involved two steps. For the first step (Step 1) we have repeated the statistical modelling which, in the September 2022 evaluation of benchmarking factors[7], looked at the shortlisted factors to help establish which of these would be prioritised to form the proposed benchmarking factors. We consider that the factors listed in Table 2 of the consultation document represent a reasonable starting point for our evaluation of the benchmarking factors to support decisions for the 2023 NSS publication. We have also considered comments from respondents to the consultation on the approach to publication of results of the National Student Survey[8] who raised concerns about specific benchmarking factors, and our responses to those comments are in our consultation response. The aim of Step 1 was to determine whether the results of statistical modelling based on shortlisted factors remain consistent with the previous review. This allows us to determine whether there is evidence to suggest that there is a reduced fit of the proposed benchmarking factors with the statistical properties described within the principles for the selection and application of benchmarking factors.

3. The second step (Step 2) aimed to determine whether the benchmarking factors and groupings we proposed in the consultation continue to achieve appropriate statistical properties. We considered analysis to support an assessment of the statistical integrity of the benchmarking approach. This step primarily considered the actual contribution to benchmark for each provider.

4. Our analysis showed consistency of the factors that were correlated across the NSS questions. So, for simplicity and brevity, the analysis presented here focused on the results from question

[6] See Proposal 5: Benchmarks at Consultation on the approach to publication of results of the National Student Survey - Office for Students.

[7] See Analysis of responses to consultation and decisions: Addendum, paragraph 43 onwards. Available at: Outcome and experience data - Office for Students.

[8] Consultation on the approach to publication of results of the National Student Survey - Office for Students

16: How well have teaching staff supported your learning? This question had the highest correlation with other questions in the survey.

Step 1

5. In Step 1 we have produced statistical models based on the factors in Table 2 of the consultation document. Factors have been included with the same groupings as were used in the table. Separate models were produced for each mode of study, and for all modes of study combined. Table A1 summarises the results of these statistical models and shows the maximum size of the estimated differences (for the attribute with the largest significant difference in each case) for each of the shortlisted factors, by measure and mode of study. It aims to provide an indication of the relative correlation with the outcome measured.

Table A1: Maximum estimated differences for the shortlisted factors, by indicator and mode of study

Key: Cells marked with grey shading identify factors not shortlisted for a particular indicator or mode. Cells showing '-' indicate no significant differences for that factor. Estimated differences in brackets correspond to populations of fewer than 5,000 students, where the difference shown in the same cell without brackets is the largest estimated difference for populations of at least 5,000. For some factors, unknown or N/A categories were included in the modelling, but these were not considered when identifying the largest estimated differences for each factor.

Maximum estimated differences (percentage points, significant at the 95 per cent level only)	Question 16			
Factor	**All**	**Full-time**	**Part-time**	**Apprentice**
Age on entry	5.3	5.8	3.2	- (7.3)
Disability	3.3	3.3	3.0	- (4.0)
Ethnicity	3.6	3.5	- (3.4)	- (6.6)
Level of study	5.8	5.7	-	- (4.4)
Mode of study	0.6			
Sex	1.1	1.2		
Subject	8.4	9.7	- (6.0)	- (5.3)

6. While Table A1 reports only the maximum size of the estimated differences per factor, there are significant nuances in the underlying data that cannot be succinctly summarised here. This includes understanding the proportion of attributes with statistically significant estimated differences and the relative population sizes across attributes.[9] Also note that none of the

[9] As described at paragraphs 29-31, and 160-163, of the January 2022 benchmarking factors review (available at Review of the selection and grouping of benchmarking factors - Office for Students), differences were estimated from statistical models with uncertainty, as indicated by 95 per cent confidence intervals. The 95 per cent significance level was primarily chosen to be illustrative of the observable statistical uncertainty. It also provides a tolerance of 'Type II' errors that suits our uses on this occasion, based on our expert judgement.

groups under the apprenticeship mode of study consisted of more than 5,000 students, and there may be difficulty getting modelling results for this group when trying to run the model with more factors.

7. We conclude, from Table A1 and from our consideration of the full results, that the results of statistical modelling based on shortlisted factors remain consistent with those previously used for the NSS. We consider that the proposed benchmarking factors continue to demonstrate strong correlation with the student positivity measures and theme measures, and that there is no evidence to suggest a reduced fit of the proposed benchmarking factors with the statistical properties described within the principles for the selection and application of benchmarking factors (as outlined in Annex E).

Step 2

8. In Step 2 we aimed to determine whether the benchmarking factors and groupings we proposed in the consultation continue to achieve appropriate statistical properties.

9. To understand how the proposed benchmarking groups impact the benchmarking process, we have calculated the contribution of providers towards their own benchmarks. This statistic 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.

10. Table A2 shows the proportion of (teaching) providers that contribute more than 5 per cent, 20 per cent, or 50 per cent towards their own benchmarks, based on the proposed benchmarking factors.

Table A2: Providers' contributions to their own benchmarks based on the proposed benchmarking factors

Percent of providers with at least (X%) contribution to their own benchmark	Question 16			
Mode	Number of providers	>5%	>20%	>50%
Full-time	415	39%	7%	1%
Part-time	108	62%	15%	5%
Apprenticeship	102	62%	9%	0%
All modes	439	44%	8%	1%

11. Based on the results shown in Table A2, we conclude that the number of unique benchmarking groups is not so large that the potential for self-benchmarking increases to unmanageable levels. We consider the risk that benchmarks calculated for these measures could become ineffective is not materially higher than it was when we made our consultation proposals and that those proposals therefore achieve an appropriate balance of policy objectives and statistical properties, as required by the principles for the selection and application of benchmarking factors.

Annex B: Brief review of benchmarking factors – question 28 (overall satisfaction)

This annex is aimed at readers seeking more information about how we have decided on the benchmarking factors that are used to produce benchmarks for NSS results.

This annex describes the brief review carried out to determine which factors to use for NSS question 28, which relates to overall satisfaction.

1. In the consultation on the approach to publication of results of the NSS[10] we proposed that we would conduct a brief review of the existing benchmarking factors. This review seeks to confirm that the statistical properties of the existing selection and grouping of benchmarking factors are not fundamentally different than those which have previously been observed in results from the 2022 and earlier NSS surveys. For the initial NSS 2023 publication, we would expect to make only minor modifications to the factors set out in Table 2 of the consultation document (such as an adjustment to the groupings used), or no changes.

2. Annex A relates to the process used to briefly review the factors used to benchmark questions 1 to 27 and the theme measures. Analysis showed that question 28 had different characteristics, and that question is discussed here. The reason for the difference is that a much smaller number of respondents answered question 28, because it was asked to students at providers in Scotland, Wales and Northern Ireland, but not those at providers in England. This means that the total number of respondents, which is the basis for creating benchmarks, is 49,000, instead of around 338,000 for the other questions.

3. For this review, our evaluation of the benchmarking factors has involved three steps. For the first step (Step 1) we have repeated the statistical modelling which, in the September 2022 evaluation of benchmarking factors[11], looked at the shortlisted factors to help establish which of these would be prioritised to form the proposed benchmarking factors. We consider that the factors listed in Table 2 of the consultation document represent a reasonable starting point for our evaluation of the benchmarking factors to support decisions for the 2023 NSS publication. We have also considered comments from respondents to the consultation on the approach to publication of results of the National Student Survey[12] who raised concerns about specific benchmarking factors, and our responses to those comments are in our consultation response. The aim of Step 1 was to determine whether the results of statistical modelling based on shortlisted factors remain consistent with the previous review. This allows us to determine whether there is evidence to suggest that there is a reduced fit of the proposed benchmarking factors with the statistical properties described within the principles for the selection and application of benchmarking factors.

[10] See Proposal 5: Benchmarks at Consultation on the approach to publication of results of the National Student Survey - Office for Students.

[11] See Analysis of responses to consultation and decisions: Addendum, paragraph 43 onwards. Available at: Outcome and experience data - Office for Students.

[12] Consultation on the approach to publication of results of the National Student Survey - Office for Students

4. The second step (Step 2) aimed to determine whether the benchmarking factors and groupings we proposed in the consultation continue to achieve appropriate statistical properties. We considered analysis to support an assessment of the statistical integrity of the benchmarking approach. This step primarily consisted of a consideration of the actual contribution to benchmark for each provider.

5. If there are concerns from Step 1 or Step 2 about the factors used, the third step (Step 3) is to explore removing factors from the list to see whether this improves the balance of policy objectives and statistical properties.

Step 1

6. In Step 1 we have produced statistical models based on the factors in Table 2 of the consultation document. Factors have been included with the same groupings as were used in the table. Separate models were produced for each mode of study, and for all modes of study combined. Table B1 summarises the results of these statistical models and shows the maximum size of the estimated differences (for the attribute with the largest significant difference in each case) for each of the shortlisted factors, by measure and mode of study. It aims to provide an indication of the relative correlation with the outcome measured.

Table B1: Maximum estimated differences for the shortlisted factors, by indicator and mode of study – Question 28

Key: Cells marked with grey shading identify factors not shortlisted for a particular indicator or mode. Cells showing '-' indicate no significant differences for that factor. Estimated differences in brackets correspond to populations of fewer than 5,000 students, where the difference shown in the same cell without brackets is the largest estimated difference for populations of at least 5,000. For some factors, unknown or N/A categories were included in the modelling, but these were not considered when identifying the largest estimated differences for each factor.

Maximum estimated differences (percentage points, significant at the 95 per cent level only)	Question 28			
Factor	**All**	**Full-time**	**Part-time**	**Apprentice**
Age on entry	(5.0)	(4.8)	(12.5)	-
Disability	3.8	3.8	-	-
Ethnicity	2.4 (3.9)	2.5 (4.0)	(38.9)	(23.2)
Level of study	(5.0)	(5.3)	-	(23.2)
Mode of study	(10.0)			
Sex	7.3	1.9		
Subject	(14.2)	13.7	(12.2)	(39.0)

7. While Table B1 reports only the maximum size of the estimated differences per factor, there are nuances in the underlying data that cannot be succinctly summarised here. This includes understanding the proportion of attributes with statistically significant estimated differences and

the relative population sizes across attributes.[13] Also note that very few of the groups consisted of more than 5,000 students.

8. We conclude, from Table B1 and from our consideration of the full results, that the results of statistical modelling based on shortlisted factors are showing some differences with those previously used for the NSS. There are some factors which are not showing as significant, and particularly for part-time and apprenticeship students, even the groups that are significant are based on small numbers of students (less than 100 students for ethnicity and level of study).

Step 2

9. In Step 2 we aimed to determine whether the benchmarking factors and groupings we proposed in the consultation continue to achieve appropriate statistical properties.

10. To understand how the proposed benchmarking groups impact the benchmarking process, we have calculated the contribution of providers towards their own benchmarks. This statistic 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. Table B2 shows the proportion of (teaching) providers that contribute more than 5 per cent, 20 per cent, or 50 per cent towards their own benchmarks, based on the proposed benchmarking factors.

Table B2: Providers' contributions to their own benchmarks based on the proposed benchmarking factors

Percent of providers with at least (X%) contribution to their own benchmark	Question 28			
Mode	Number of providers	>5%	>20%	>50%
Full-time	46	98%	39%	0%
Part-time	20	100%	85%	5%
Apprenticeship	18	100%	72%	17%
All modes	47	98%	43%	0%

11. Based on the results shown in Table B2, we conclude that:

    a. For full-time courses, the number of unique benchmarking groups is not large that the potential for self-benchmarking increases to unmanageable levels. We consider the risk that benchmarks calculated for these measures could become ineffective is not materially

[13] As described at paragraphs 29 to 31, and 160 to 163, of the January 2022 benchmarking factors review (available at Review of the selection and grouping of benchmarking factors - Office for Students), differences were estimated from statistical models with uncertainty, as indicated by 95 per cent confidence intervals. The 95 per cent significance level was primarily chosen to be illustrative of the observable statistical uncertainty. It also provides a tolerance of 'Type II' errors that suits our uses on this occasion, based on our expert judgement.

higher than it is for the other questions, and given the smaller number of providers and students, is expected that we will have to tolerate a higher level of contributions to benchmark. For full-time courses, these factors therefore achieve an appropriate balance of policy objectives and statistical properties, as required by the principles for the selection and application of benchmarking factors.

b. For part-time and apprenticeship courses, there are concerns that the potential for self-benchmarking is increasing to a level where it is detrimental to the use of the data. The risk that these benchmarks become ineffective is higher than for the other questions in the survey. This means it is necessary to explore whether a smaller number of factors would achieve a more appropriate balance of policy objectives and statistical properties, as required by the principles for the selection and application of benchmarking factors.

Step 3

12. Therefore, since we have concerns about some of the factors used for part-time and apprenticeship students, in Step 3 we will consider removing some factors and see the statistical effect, before weighing the options against policy objectives.

13. There are five factors used for benchmarking part-time and apprenticeship students: age on entry, disability, ethnicity, level of study and subject of study. Looking at Table B1:

    a. Age on entry is significant for part-time students, and is quite close to being significant for both age group categories for apprenticeship students.

    b. Disability is not significant for either part-time or apprenticeship students.

    c. Ethnicity appears as significant in the table, but the ethnic groups which are significant are very small – in both cases it is the small ‘other ethnic groups’ category which has the highest difference.

    d. Level of study and subject of study are both categories users can split by in the published results, and benchmarks are produced for each of these splits. Even if we removed them from the benchmarking factors, the benchmarking groups would often still be divided into these splits. Leaving them as factors is the simplest way to ensure that the statistical properties of the factors are robust to displaying results by level and subject of study. In addition, subject is significant for both modes of study, and level of study is significant for apprenticeship students (albeit with very small numbers).

14. From this analysis, the preferable option for the short term would be to remove disability and ethnicity as benchmarking factors, but keep age on entry, level of study and subject of study. Table B3 shows the contribution of providers towards their own benchmark if we use these factors for part-time and apprenticeship students.

Table B3: Provider's contribution to their own benchmarks based on the modified benchmarking factors (without disability or ethnicity)

Percent of providers with at least (X%) contribution to their own benchmark	Question 28			
Mode	Number of providers	>5%	>20%	>50%
Full-time	46	98%	39%	0%
Part-time	20	100%	65%	0%
Apprenticeship	18	100%	50%	6%
All modes	47	98%	40%	0%

15. This shows that there are still some concerns about the level of self-benchmarking, but reducing the number of benchmarking factors has been an improvement. In particular, there are now no providers with part-time contributions to benchmark over 50 per cent and much fewer for apprenticeship courses. Some high contributions to benchmark are inevitable given the smaller number of providers with part-time and apprenticeship courses in Wales, Scotland and Northern Ireland. This is therefore the option we will proceed with for the NSS 2023 publication, as it achieves the best balance of policy objectives and statistical properties. In the longer term, we intend to seek further evidence that the inclusion of age on entry in the benchmarking of part-time and apprenticeship students does not represent an arbitrary selection from the wider range of student characteristics used in benchmarking the full-time students.

16. To conclude, for full-time students we will continue to use the same factors for question 28 as for the other questions: mode, level and subject of study, along with sex, age on entry, disability and ethnic group. For part-time and apprenticeship students we will use fewer factors for question 28 than for the other questions: mode, level and subject of study, and age on entry. We will conduct a fuller review of benchmarking factors to inform the benchmarking factors that will be used in the NSS 2024 and later publications.

Annex C: Worked example of benchmarking calculations

This annex is aimed at readers seeking more information about how we have calculated the benchmarks that are reported within the NSS results.

This annex includes a fictional, simplified example to demonstrate how we calculate benchmarks for a single question. This example demonstrates the method that applies to the calculation of benchmarks for all NSS questions and theme measures.

1. In this fictional, simplified example, assume that we are seeking to calculate benchmarks for survey question 1 using only two benchmarking factors which affect the positivity rate for that question. Specifically, we want to take account of students' age on entry to higher education, and the subject that they are studying. Suppose that students' age is defined as either 'young' or 'not young' and that the higher education sector delivers provision in only three subject areas (agriculture, maths and history). The figures given in this annex are for illustrative purposes only and are not reflective of provider or sector behaviour.

2. That means that for this measure there are six possible distinct benchmarking groups, set out in the table below.

Step one: the provider

3. The provider for which we are calculating a benchmark has 1,090 students studying agriculture and maths. Table C1 shows the provider's students, split across the six benchmarking groups, and the positivity measure that we observe for each of these groups.

4. Overall, the provider has a positivity rate of 94.3 per cent. This is effectively a weighted average of the rates for each group.

5. Note that the provider's observed positivity measure for young maths students is particularly low (92.0 per cent) in comparison to the observed measure for other groups at the provider. This low positivity measure is outweighed by the larger number of students in groups with higher observed positivity measures, such as young agriculture students.

Table C1: Distribution of the provider's observed positivity measures across benchmarking groups

Age group	Subject group	Number of students	Students in the benchmarking group as a proportion of total students	Observed positivity measure
Young	Agriculture	500	45.9%	95.0%
Young	History	0	0.0%	N/A
Young	Maths	150	13.8%	92.0%

Age group	Subject group	Number of students	Students in the benchmarking group as a proportion of total students	Observed positivity measure
Not young	Agriculture	400	36.7%	94.0%
Not young	History	0	0.0%	N/A
Not young	Maths	40	3.7%	98.0%
				Provider positivity measure
Total		1,090	100%	94.3%

Step two: the sector

6. There are 210,500 full-time students across the whole sector, studying agriculture, maths and history. Table C2 shows the sector's students, split across the six benchmarking groups, and the positivity measure that we observe for each of these groups across the sector as a whole.

7. Overall, the sector has a positivity measure of 96.6 per cent.

8. Note that the sector's overall positivity measure is driven by high positivity measures observed for young history students (99.0 per cent), and the small student numbers for agriculture subjects, for which we observe relatively low rates for both young (95.0 per cent) and not young (94.0 per cent) students.

Table C2: Distribution of the sector's observed positivity measures across benchmarking groups

Age group	Subject group	Number of students	Observed positivity measure
Young	Agriculture	20,000	95.0%
Young	History	80,000	99.0%
Young	Maths	95,000	95.0%
Not young	Agriculture	5,000	94.0%
Not young	History	6,500	98.0%
Not young	Maths	4,000	98.0%
			Sector positivity measure
Total		210,500	96.6%

Step three: calculating the provider specific benchmark

9. So far, in Table C2, the sector's positivity rates are weighted against the numbers of students in the **sector** in each of the six distinct benchmarking groups. In Table C3 below, the sector's positivity measures are instead weighted to reflect the students in the **provider**.

10. Table C3 shows that weighting the sector's positivity measures by the proportion of students in each benchmarking group at the provider results in a weighted sector benchmark of 94.7 per cent for this provider.

11. This weighted sector rate is lower than the original sector rate shown in Table C2 since it no longer reflects the (relatively high) rates for history students (because the provider has no history students), and because the agriculture groups have a much higher weighting, reflecting that the provider has a higher proportion of agriculture students than the sector as a whole.

12. The provider's overall positivity measure (94.3 per cent) can now be compared with the weighted sector benchmark (94.7 per cent). The provider's positivity measure is still lower than the rate observed for students with similar characteristics across the sector.

Table C3: Calculation of the provider benchmark using the sector's observed continuation rates across benchmarking groups

Age group	Subject group	Students in the benchmarking group as a proportion of total students at the provider (a)	Sector observed positivity measure (b)	Weighted sector positivity numbers (= a x b)
Young	Agriculture	45.9%	95.0%	43.6%
Young	History	0.0%	99.0%	0.0%
Young	Maths	13.8%	95.0%	13.1%
Not young	Agriculture	36.7%	94.0%	34.5%
Not young	History	0.0%	98.0%	0.0%
Not young	Maths	3.7%	98.0%	3.6%
Total		100%	Sector overall measure	**Provider benchmark**
			96.6%	**94.7%**
				(= 43.6% + 0.0% + 13.1% + 34.5% + 0.0% + 3.6%)

Annex D: Definitions of subject areas of study groupings used in benchmarking

This annex is aimed at readers seeking more information about how we have calculated the benchmarks that are reported within the NSS results.

This annex includes definitions of the benchmarking factors of subject areas of study, which are used to benchmark all NSS questions and theme measures.

1. Table D1 shows the groupings of subject areas of study that we have decided to use as benchmarking factors. We have decided to use these groupings as follows:

   - Broadly defined subject groups as benchmarking factors for part-time and apprenticeship students.
   - CAH level 2 groups as benchmarking factors for full-time students.

Table D1: Groupings of subject areas used as benchmarking factors

Broadly defined subject group	CAH level 2 group
Medicine, dentistry and veterinary sciences	CAH01-01: Medicine and dentistry
	CAH05-01: Veterinary sciences
Nursing, allied health and psychology	CAH02-02: Pharmacology, toxicology and pharmacy
	CAH02-04: Nursing and midwifery
	CAH02-05: Medical sciences
	CAH02-06: Allied health
	CAH04-01: Psychology
Natural and mathematical sciences	CAH03-01: Biosciences
	CAH03-02: Sport and exercise sciences
	CAH07-01: Physics and astronomy
	CAH07-02: Chemistry
	CAH07-04: General, applied and forensic sciences
	CAH09-01: Mathematical sciences
Engineering, technology and computing	CAH10-01: Engineering
	CAH10-03: Materials and technology
	CAH11-01: Computing
Law and social sciences	CAH15-01: Sociology, social policy and anthropology

Broadly defined subject group	CAH level 2 group
	CAH15-02: Economics
	CAH15-03: Politics
	CAH15-04: Health and social care
	CAH16-01: Law
Business and management	CAH17-01: Business and management
Humanities and languages	CAH19-01: English studies
	CAH19-04, CAH19-02: Languages and area studies
	CAH20-01: History and archaeology
	CAH20-02: Philosophy and religious studies
	CAH23-01: Combined and general studies
	CAH24-01: Media, journalism and communications
Education and teaching	CAH22-01: Education and teaching
Design, and creative and performing arts	CAH25-01: Creative arts and design
	CAH25-02: Performing arts
Natural and built environment	CAH06-01: Agriculture, food and related studies
	CAH13-01: Architecture, building and planning
	CAH26-01: Geography, earth and environmental studies

Annex E: Principles for the selection and application of benchmarks

This annex is aimed at readers seeking more information about how we have selected benchmarking factors. The benchmarking principles outlined in this annex have guided the selection and application of benchmarking factors for benchmarking indicators and split indicators for the NSS results.

1. These principles will be guiding rather than binding, but they are intended to provide an effective mechanism to build public trust and confidence in the benchmarks that the OfS creates and uses in its student outcome and experience indicators.

2. When selecting benchmarking factors, the intention is that each principle would be considered in turn and, where appropriate, evidence of its applicability would be sought from statistical analysis or modelling. We are aware that the principles may sometimes sit in tension with one another, and that in most cases a value-based judgement will be required to confirm fit or applicability with the principle.

3. The proposed **core principles** relating to the factors being considered for benchmarking are:

    a. The selection of benchmarking factors should be fit for purpose, evidence-based and robust, conforming to recognised best practice in the production of statistical information. In particular:

        i. Details of the selection process should be published for the benefit of providers and other users or interested parties.

        ii. The selection of benchmarking factors should vary across different student outcome and experience indicators only when there is a clear and valid rationale.

        iii. The number and definition of benchmarking factors selected should not compromise the statistical integrity of the broader benchmarking approach.

    b. Benchmarking factors should be applicable to, and available for, all types of providers across England that are delivering the higher education provision for which the indicator is measuring students' outcomes or experience.

    c. Benchmarking factors should contribute to an overall benchmarking approach which supports fair comparison of indicators across the higher education sector. A candidate benchmarking factor should therefore have relevance to help explain the context or differing characteristics of a provider's students or provision.

    d. The benchmarking approach should neutralise the effect of characteristics on a provider's performance where this is consistent with policy objectives. This approach guards against inadvertently creating incentives for providers to change their behaviour in terms of the students they recruit or the range of provision they offer in ways that could undermine our ability to meet our duties around access and participation, and competition. It does not imply that it is acceptable for some student groups to receive lower quality provision, but

recognises that this is currently the case and the risks of not controlling for it. The benchmarking approach should only neutralise the effect of characteristics where there is such a risk of negative unintended consequences, as otherwise it risks creating perverse incentives.

e. Benchmarking factors should primarily reflect structural factors that contribute to variations in student outcomes or experience which are outside a provider's control, or undesirable for it to control for. This means that characteristics of the provider will not normally act as benchmarking factors.

f. In selecting the range of benchmarking factors to apply for a given indicator, the need to preserve the statistical integrity of the broader benchmarking approach requires that consideration should be given to limit the number of factors on the basis of:

   i. The size of the population for which the effect occurs: it is unlikely that a factor where the effect is limited to a small population will be selected where there are other factors with similar effects that have broad applicability.

   ii. The distribution of the population for which the effect occurs: it is unlikely that a factor where the effect is limited to a population concentrated in a small subsection of providers will be selected where there are other factors with similar effects that have applicability to a wider cross-section of provision.

   iii. The nature of the other candidate factors: where there are a number of similar candidate factors (for example, measures of disadvantage), it will normally be the case that only the one that has the greatest effect should be selected so that a balance of factors is achieved.

g. The factors used in benchmarking should be reviewed at regular intervals, to check that the evidence for, and applicability of, the approach remains current and fit for purpose, and to consider the impact achieved by previous benchmarking exercises.

Availability and data quality

4. The availability and data quality of candidate benchmarking factors should be considered in relation to the principles as follows:

   a. The quality of data items considered as candidate benchmarking factors should be understood and judged to be of sufficiently high quality for use in a benchmarking exercise. The data items should normally be collected in a consistent and fair way across the sector; it should have a good sample base and use transparent definitions.

   b. Where possible, benchmarking factors should be drawn from existing data sources. Any proposal to collect further data for the purpose of a benchmarking factor should be carefully considered against the principles for data burden

Statistical properties

5. The principles for the statistical properties that candidate benchmarking factors should demonstrate are:

a. Statistical models that seek to account for a range of characteristics should identify a remaining correlation between the benchmarking factor and the student outcome or experience that is being measured.

b. Once other factors have been accounted for, statistical modelling should identify that the performance being measured is not uniformly distributed across the attributes within a benchmarking factor, and that differences between these attributes are non-trivial.

c. A benchmarking factor should not be uniformly distributed across providers or performance units; rather, the factor should differentially affect the benchmarks that are calculated, meaning that factors which are distributed unevenly across providers or performance units should be considered as stronger candidates to be used as benchmarking factors.

d. Where possible, a benchmarking factor should be a direct measure, rather than a proxy.

e. As far as possible, the selection of benchmarking factors should limit the extent to which a benchmark value can be determined by a single provider. The selection of a benchmarking factor (and the subsequent grouping of attributes within it) should not compromise the statistical integrity of the broader benchmarking approach.

f. Benchmarking factors (and the data sources from which they are derived) should normally have longevity, with these statistical properties observed to continue over time.

Defining groupings of the attributes within the benchmarking factor

6. Once benchmarking factors have been selected, the principles for defining groupings of the attributes within the benchmarking factor are:

    a. The grouping of attributes within benchmarking factors should be fit for purpose and determined through consideration of sound evidence.

    b. The number of categories formed when grouping attributes within benchmarking factors should be the minimum for the benchmarking factor to be effective. The number and definition of the groupings should not compromise the statistical integrity of the broader benchmarking approach.

    c. The grouping of attributes within benchmarking factors should avoid creating groups in which numbers of students possessing those attributes are either very small or very large in the sector overall. The effect of creating groups that are known to be very small or very large at individual provider level should be acknowledged where they cannot be avoided.

    d. The attributes that form a grouping should share a consistency of student backgrounds, outcomes or behaviours with respect to the indicator to which they refer. The consistency of attributes should be understood from the evidence of statistical analysis.

    e. The grouping of attributes within benchmarking factors should make practical sense, to form coherent groups which share a qualitative similarity.

f. The grouping of attributes within benchmarking factors should vary across indicators only when there is a clear and valid rationale. Where variations are necessary, those deviations should use other groupings that exist elsewhere in a sector-wide hierarchical view of the benchmarking factor in question, at a more aggregated or disaggregated level according to need.

g. The grouping of attributes within benchmarking factors should be reviewed periodically to ensure that it continues to comply with these principles.

Annex F: Technical detail about benchmarking calculations

This annex is aimed at readers seeking to understand the calculation of benchmark values from individualised student data. It provides information about the calculation of benchmarks.

The information in the annex is aimed at readers with an in-depth knowledge of advanced statistical methods and assumes a familiarity with statistical formulae and notation.

1. The general approach to benchmarking follows the design-based adjustment method described in 'Statistical analysis of performance indicators in UK higher education' by Draper and Gittoes (2004).[14] This annex summarises the key information from that methodology.

General approach

2. In this method, for each unique combination of benchmarking factors (described as potential confounding factors (PCFs) in the literature), an observed rate for the measure, and the number of students that inform it, is calculated for both the sector and each provider.

3. The presentation of these rates and number of students for each unique combination of benchmarking factors can be visualised as two large grids as shown in Figure H1 below (the rates shown in the top table, with the number of students in the bottom table). In this figure, M represents the number of unique combinations of benchmarking factors. The method is based on a further cross-tabulation of the N providers by these M categories. The '.' and '+' notations in subscripts indicate averaging and summing over the relevant columns or rows of the table respectively. Within each table, each cell ij contains n_{ij} students from provider i with unique combination of benchmarking factors j. The observed rate of success of these students is $\hat{p}_{ij}$. Each weighted row mean, $\hat{p}_{i.}$ is the observed indicator value for provider i and $\hat{p}_{.j}$ is the observed indicator value for students with unique combination of benchmarking factor j across all students in the sector.

[14] Draper, D and Gittoes, M (2004). Statistical analysis of performance indicators in UK higher education. Journal of the Royal Statistical Society. Series A (Statistics in Society), 167, Part 3, pages 449-474.

Figure F1: A tabular presentation of the rates and number of students for each unique combination of benchmarking factors per provider

Provider	Unique combination of benchmarking factors 1	2	…	M	Weighted row mean
1	$\hat{p}_{11}$	$\hat{p}_{12}$	…	$\hat{p}_{1M}$	$\hat{p}_{1\cdot}$
2	$\hat{p}_{21}$	$\hat{p}_{22}$	…	$\hat{p}_{2M}$	$\hat{p}_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
N	$\hat{p}_{N1}$	$\hat{p}_{N2}$	…	$\hat{p}_{NM}$	$\hat{p}_{N\cdot}$
Weighted column mean	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	…	$\hat{p}_{\cdot M}$	$\hat{p}_{\cdot\cdot}$

					Row sum
1	n_{11}	n_{12}	…	n_{1M}	n_{1+}
2	n_{21}	n_{22}	…	n_{2M}	n_{2+}
⋮	⋮	⋮	⋱	⋮	⋮
N	n_{N1}	n_{N2}	…	n_{NM}	n_{N+}
Column sum	n_{+1}	n_{+2}	…	n_{+M}	n_{++}

4. The observed indicator value, $\hat{p}_{i\cdot}$, for the provider can be directly read from the tables in Figure F1. The structure of the table allows us to consider the question: 'What would the observed indicator value have been at provider i, if its distribution of students across the unique combination of benchmarking factors had been what is was, but its rates were replaced by the sector rates, $\hat{p}_{\cdot j}$?'. These can be summarised as follows:

   The observed indicator value, $\hat{O}_i$, at provider i is:

$$\hat{O}_i = \hat{p}_{i\cdot} = \frac{1}{n_{i+}} \sum_{j=1}^{M} n_{ij} \hat{p}_{ij}$$

   The benchmark, $\hat{E}_i$, at provider i is:

$$\hat{E}_i = \frac{1}{n_{i+}} \sum_{j=1}^{M} n_{ij} \hat{p}_{\cdot j}$$

   The difference between the observed indicator value and benchmark, $\hat{D}_i$, at provider i is:

$$\hat{D}_i = \hat{O}_i - \hat{E}_i$$

5. To aid interpretation of the observed difference, the standard deviations of the differences between the indicator value and benchmark have been calculated. A standard deviation measures the amount by which one would expect a statistic to change, based solely on random sampling. Because these are standard deviations of a statistic (the difference), they are more usually called standard errors.

6. To calculate the standard deviation, the formula for the difference is adjusted using algebraic manipulation (the full manipulation can be found in the literature) to be written as a weighted sum of all cells in the tables shown in Figure F1:

$$\widehat{D}_i = \sum_{j=1}^{M}\sum_{k=1}^{N} \lambda_{ikj}\hat{p}_{kj}$$

where $\lambda_{ikj} = \frac{n_{ij}}{n_{i+}}(\delta_{ik} - \frac{n_{kj}}{n_{+j}})$

and $\delta_{ik} = 1 \quad if\ i = k,$
$\delta_{ik} = 0 \quad if\ i \neq k$

Assuming the $\hat{p}_{kj}$ terms are independent, the variance is given by:

$$Var(\widehat{D}_i) = \sum_{j=1}^{M}\sum_{k=1}^{N} \lambda_{ikj}^2 Var(\hat{p}_{kj})$$

The literature shows that a reasonable estimate for the variance of $\hat{p}_{kj}$ can be made by using a shrinkage estimation procedure:

$$Var(\hat{p}_{kj}) = \frac{\hat{p}^*_{kj}(1-\hat{p}^*_{kj})}{n_{kj}}$$

where $\hat{p}^*_{kj} = 0.5\hat{p}_{..} + 0.5\hat{p}_{kj}$

and $\hat{p}_{..}$ is the overall rate of the sector.

The square root of the variance of $\widehat{D}_i$ gives the standard deviation.

7. We calculate the average contribution to benchmark for provider, i, using a similar weighted average calculation. This statistic calculates the contribution of the provider's own students on the sector averages that informs the calculation of the provider's benchmark of the form:

$$average\ contribution\ to\ the\ benchmark_i = \sum_{j=1}^{M} \frac{n_{ij}^2}{n_{+j}n_{i+}}$$

Benchmarking split indicators

8. In the calculation of the standard deviation for the purposes of benchmarking split indicators a small adjustment is made within the formulae described in the general approach above. The

approach to create an estimate for the variance of $\hat{p}_{kj}$ by using a shrinkage estimation is the same, but the value for $\hat{p}_{..}$ used in the derivation of $\hat{p}^{*}_{kj}$ remains the overall rate of the sector calculated at provider level. This is instead of using $\hat{p}_{..}$ created based on the subset of the provider and sector to the split indicator. This adjustment is made to ensure that the shrinkage estimation is applied consistently between the overall provider split indicator and other split indicators. For example, in a case where a provider delivers only a single subject, the standard deviation could appear different for the provider-level indicator and the split for the subject only because of the shrinkage estimation.

9. These differences in the approach to calculating benchmarks for split indicators is presented in the same tabular presentation as in Figure F1 in Figure F2, which assumes the split indicator being calculated is for 'Male' students. The $\hat{p}_{..}$ has been relabelled as $\widehat{Overallp}_{..}$. Otherwise, the notation is the same as described in paragraph three of this annex.

Figure F2: A tabular presentation of the rates and number of students for each unique combination of benchmarking factors per provider for male students

Male students at provider...	Unique combination of benchmarking factors 1	2	...	M	Weighted row mean
1	$\hat{p}_{11}$	$\hat{p}_{12}$	...	$\hat{p}_{1M}$	$\hat{p}_{1\cdot}$
2	$\hat{p}_{21}$	$\hat{p}_{22}$	...	$\hat{p}_{2M}$	$\hat{p}_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
N	$\hat{p}_{N1}$	$\hat{p}_{N2}$	...	$\hat{p}_{NM}$	$\hat{p}_{N\cdot}$
Weighted column mean	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	...	$\hat{p}_{\cdot M}$	$\hat{Overallp}_{..}$

					Row sum
1	n_{11}	n_{12}	...	n_{1M}	n_{1+}
2	n_{21}	n_{22}	...	n_{2M}	n_{2+}
⋮	⋮	⋮	⋱	⋮	⋮
N	n_{N1}	n_{N2}	...	n_{NM}	n_{N+}
Column sum	n_{+1}	n_{+2}	...	n_{+M}	n_{++}

10. Note that this approach will be modified slightly for the CAH3 splits, where all courses in the CAH2 subject group are included in the formation of the benchmark. The formula to calculate the average contribution to benchmark[15] will also be based on the CAH2 subject group. See paragraphs 22-29 for more information.

[15] See paragraph seven in this annex.