

## Methodology document for National Student Survey sector analysis

1. This document explains the methodology used to produce the sector-level analysis of National Student Survey (NSS) data published on 15 July 2021.

### Background and scope

2. This data provides an overview of responses to the 2021 NSS and explores the extent to which responses differ across student groups.
3. The data includes NSS results from the 2020 and 2021 surveys. Respondents are included from all domiciles and all levels of study. The analysis includes students studying in England, Northern Ireland, Scotland and Wales. For more information about which students are eligible to participate in the survey, see the NSS good practice guide.<sup>1</sup> The analysis is at sector level, so no provider level results are included.
4. The data is presented in three interactive charts:
  - a. The first, 'Sector results', displays display the percentage of respondents who agreed (selected 'Definitely agree' or 'Mostly agree') – known as the agreement rate – for all questions and scales<sup>2</sup> side by side, with the ability to filter results by nation and mode of study. This chart is designed to provide an overall picture of students' opinions on the areas covered by the NSS.
  - b. The second and third, 'Results by characteristic' and 'Differences by characteristic', display the agreement rate for each question and scale in the NSS. Responses have been separated according to six student and course characteristics: age, sex, ethnicity, whether the student reported having a disability, subject classification and mode of study. Alongside the actual percentage of respondents who agreed to each question, benchmark values have been calculated using these characteristics. We have shown when the agreement rate for a characteristic group differs from the benchmark.
5. There is always a degree of uncertainty around this sort of measurement. For example, the NSS survey is a measurement at a point in time – it is possible that some respondents would respond differently on a different day. To account for this, we have indicated when the difference from the benchmark is likely to be significant, rather than due to this sort of random variation. Where benchmarks are not available (i.e. the 'Sector results' chart), we have calculated the 95% confidence interval around the agreement rate.
6. A full description of the characteristics used in this analysis and their values is shown in [Table 3](#). The data is shown separately for UK-domiciled and non-UK domiciled students. Categories

---

<sup>1</sup> See pp. 7-8 of the NSS 2021 Good Practice Guide, available at [www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/promotion-of-the-nss/](http://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/promotion-of-the-nss/)

<sup>2</sup> A scale is a group of questions addressing a common theme, for example 'Assessment and feedback'.

will not be shown where the response rate is lower than 50 per cent or there are fewer than 10 responses. No categories failed to meet these publication criteria in the first publication of the data on 15 July 2021.

## Benchmarking and significance testing

7. In the 'Results by characteristic' and 'Differences by characteristic' charts, we have used benchmarking to take into account factors that are known to influence NSS agreement rates. Benchmarks are weighted sector averages. They give information about the agreement rates that might be expected given the student and course characteristics included in the weighting. Where differences exist between agreement rates and their benchmarks, this suggests that the agreement rates are affected by a factor that is not included in the weighting.
8. The benchmarks are based on the six factors standardly used to analyse NSS data, with five of these used to calculate each individual benchmark. For example, if the characteristic analysed is age, then the benchmark is based on the sex, ethnicity, whether the student reported having a disability, subject classification and mode of study of respondents.<sup>3</sup> If the agreement rate for an age group is different from the benchmark, we can infer that this difference is unlikely be due to any of the characteristics included in the benchmark: for instance, we can ignore the fact that older students are more likely to study part-time. The difference may instead be due to the age of the students, or to a further unknown factor.
9. A full explanation of the benchmarking methodology used here and in other higher education statistics can be found on the Higher Education Statistics Agency's (HESA's) website.<sup>4</sup>
10. As part of this analysis, we have evaluated the extent to which a user can be confident that the agreement rate differs from the benchmark. This is equivalent to evaluating the probability that the difference has arisen due to random variation in the responses: a user has 95 per cent confidence if the probability that the difference is due to chance is less than or equal to 5 per cent. We have presented this evaluation as follows:
  - a. By default, differences with a confidence level below 95 per cent are marked as not significant and displayed in grey.
  - b. By default, differences with a confidence level of at least 95 per cent are displayed in a colour that indicates whether the agreement rate is significantly above or below the benchmark (shown in green and purple respectively).
  - c. Users can change the confidence level from 95 per cent to any value between 50 and 99 per cent. The colours of the bars will automatically update to reflect the chosen confidence level. The maximum confidence level met by a bar can be seen by hovering the mouse over it.

---

<sup>3</sup> If the characteristic being analysed were to be included as a benchmarking factor, there would be no difference between the actual agreement rates and the benchmarks.

<sup>4</sup> See 'Technical notes': <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks>. Please note that the method described for calculating benchmarks and standard deviations has been used for this analysis, but not the method for producing significance markers; our approach to marking significance is described below. Further note that the technical notes refer to provider-level benchmarks, whereas benchmarks have been calculated at sector level only in this analysis.

This approach was chosen to reduce the ‘cliff edge’ effect produced by testing against a single, fixed confidence level, and to reflect the fact that different users may be willing to tolerate different levels of uncertainty.<sup>5</sup> The Office for Students (OfS) is currently exploring different ways of showing statistical significance and we welcome feedback on the approaches taken here.<sup>6</sup> Users of the data should be aware that in the majority of cases, changing the confidence level will not change the outcome of significance tests **in this analysis**. This is because the analysis is at sector-level, meaning the samples are often large and allow us to have a high degree of confidence in both the agreement rates and benchmarks. This means any sizeable difference between the two is often likely to be significant with a high degree of confidence.

11. It is expected that users of the data will wish look at multiple data items, and to consider together several or many comparisons with the benchmark. For example, when viewing responses split by subject classification, users may consider the agreement rates per subject as a whole, noting cases in which the rate is significantly better or worse than the benchmark. In this scenario, standard statistical tests can overstate confidence, as the probability of a false positive increases along with the number of data items under consideration.<sup>7</sup> To overcome this, an adjustment is often made to the calculation to control the false discovery rate (Benjamini and Yekutieli, 2001)<sup>8</sup>: in this case, a Bonferroni correction has been used.
12. We have considered the total number of comparisons as the product of the total number of questions and scales (35), the maximum number of categories within a single characteristic (21, for subject of study), and the number of years of data (2). This gives the total number of comparisons as 1,470. We have considered the number of comparisons within each characteristic and broad domicile group (i.e. UK or non-UK domiciled) as the total number of comparisons; this approach is cautious but stops short of assuming that users will consider all possible comparisons together. For simplicity we have assumed that all comparisons are independent and chosen to use the same correction for all characteristics. This is a conservative approach, which means that for most users the level of confidence will be understated. The OfS is reviewing its approach to dealing with multiple comparisons, and this approach may be refined in future. A selection of corrected significance levels calculated using the Bonferroni method<sup>9</sup> are shown in Table 4.
13. Benchmarks are not available for the ‘Sector results’ chart. To provide a measure of uncertainty, we have instead calculated 95% confidence intervals<sup>10</sup> around them. As mentioned in paragraph 5, the NSS is a measurement at a point in time that estimates ‘true’ underlying value in the population. Naturally, there is uncertainty in how well this measurement estimates

---

<sup>5</sup> It is common to use 95 per cent as a threshold for confidence: differences that meet this threshold would be marked as significant, and differences beneath this threshold would be marked as not significant. This can be misleading, as it exaggerates the disparity between a difference with 94 per cent confidence and one with 96 per cent confidence. Some users may consider that a difference with 94 per cent confidence is still of interest, but this information would be lost when using a fixed threshold of 95 per cent.

<sup>6</sup> Feedback should be directed to [NSS@officeforstudents.org.uk](mailto:NSS@officeforstudents.org.uk).

<sup>7</sup> For example, if 20 comparisons are made at a 95 per cent confidence level, the probability of at least one comparison resulting in a false positive is  $1 - 0.95^{20} \approx 64.2$  per cent.

<sup>8</sup> Benjamini, Yoav; Yekutieli, Daniel. ‘The control of the false discovery rate in multiple testing under dependency.’ *Ann. Statist.* 29 (2001), no. 4, 1165--1188. doi:10.1214/aos/1013699998

<sup>9</sup> The formula for corrected confidence levels using this method is given by  $1 - ((1 - C) / 2N)$  where  $C$  is the confidence level expressed as a decimal (e.g. 0.95), and  $N$  is the total number of comparisons (i.e. 1,470).

<sup>10</sup> Confidence intervals were calculated using the Wilson method. Further information about the Wilson method is available at <https://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/questions-about-the-nss-data/> under the question “What are the ‘confidence intervals’ attributed to the data?”.

the true value, and the confidence interval tells us how confident we can be that it does. It does this by providing a set of limits, between which we can be 95% confident that the true value lies. If there is no overlap between the confidence intervals for two questions in the NSS, then we can be reasonably sure that there is a genuine difference in the agreement rates for those two questions. To make sure that we do not overstate our confidence in this, we need to make a similar adjustment for multiple comparisons as we do for the significance tests, as users are likely to compare the agreement rates for multiple questions together. The same Bonferroni method was used as is described in paragraph 11, but with a different total number of comparisons. We have considered the total number of comparisons as the total number of questions and scales (35), the number of categories of mode of study (3), the number of nations (5), and the number of years of data (2). This gives the total number of comparisons as 1,050.

## Quality profile

14. With any statistical output, there are factors that influence its reliability and accuracy. We have assessed some of the factors impacting this analysis, which are explained in this section. We welcome feedback on how the reliability and accuracy of this statistic can be improved.
15. This analysis is produced using NSS data. This data is linked to the Higher Education Statistics Agency (HESA) Student and Alternative Provider (AP) Student records, and the Individualised Learner Record (collected by the Education and Skills Funding Agency); all three are established data collections with their own quality assurance mechanisms. We therefore have a high level of confidence in the source data and the subsequent processing. Although errors sometimes occur in these data sources, we judge it unlikely that an undiscovered error in a provider's data would materially affect this analysis.
16. In addition to the usual quality checks we carry out on the NSS data, for the 2021 results we examined the impact of the coronavirus pandemic on the quality of the data.<sup>11</sup> We could not identify changes in response trends that could be attributed to events during the survey window, such as the gradual easing of lockdown. While response rates vary between different student groups, this variation is no higher than in previous years. The analysis we carried out as part of the quality checks can be viewed on our website<sup>12</sup>.
17. The benchmarking method used in this analysis is an established method used by HESA and the OfS in a range of contexts, including the calculation of provider benchmarks for the routine NSS data release. We have chosen to minimise risk by using the same method, and only varying the factors we use in the benchmarks.
18. We have identified the following limitations with our approach:
  - a. The benchmarking factors used have been identified in prior reviews as being likely to influence responses to the NSS. There may be other factors that influence responses, and which may better explain differences from the benchmark in some cases. For example, if students with certain characteristics are clustered at certain providers, and these providers offer a distinctive experience, then our analysis may show that this group of students have

---

<sup>11</sup> See [www.officeforstudents.org.uk/publications/nss-2020-analysis-of-impact-of-coronavirus/](http://www.officeforstudents.org.uk/publications/nss-2020-analysis-of-impact-of-coronavirus/).

<sup>12</sup> See <http://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/nss-data-additional-analysis/>

responses that differ from the sector as a whole. Yet, in this case, it would be a mistake to infer that the students' characteristics explain or cause their different responses, as this difference may be better explained by the provider they attend.

- b. When carrying out quality checks on the 2021 NSS data, we found that whether a student is a distance learner or not had a stronger influence on how they responded to the survey this year than in previous years. Distance learning was not included as a benchmarking factor, however there is significant overlap between distance learning and being a part-time or mature learner, both of which are included in the benchmarking factors. It is possible that distance learning may therefore explain a portion of any significant differences between agreement rates and benchmarks for part-time and mature learners.
- c. The NSS has a high response rate. Nonetheless, the fact that some students do not respond introduces the possibility of non-response bias. This would occur if the non-respondents are different from the respondents in ways that systematically affect student or course characteristics.<sup>13</sup> In this analysis we have guarded against non-response bias by suppressing results for characteristic groups with response rates below 50 per cent (if any exist), and checking that response rates do not differ greatly across characteristic groups. The OfS is presently exploring other ways to identify and correct for non-response bias.
- d. The significance tests are dependent on the population size: when there are many students in a characteristic group, it is easier to be confident that a difference from the benchmark is significant. Correspondingly, we are less likely to find significant differences for characteristic groups that include fewer students. Users should be aware that a lack of a significance marker means that the data does not allow us to conclude with confidence that the difference is real. It does not imply that we can be confident that there is no difference.
- e. This analysis looks at students grouped by a single characteristic: age, sex, subject, mode, ethnicity or whether the student reported having a disability. We acknowledge that this ignores the interplay between these characteristics. For example, it is possible that the responses of part-time female students differ greatly from the responses of male part-time students; this is not explored here. The OfS has carried out work analysing outcomes by multiple characteristics, and continues to explore ways of analysing and displaying multiple characteristics. More information is available in our report 'Update to associations between characteristics of students'.<sup>14</sup>
- f. Data about the characteristics of non-UK domiciled students is not always available. This means that particular caution is needed in interpreting the results for non-UK domiciled students, as differences may be hidden or created by the lack of data. This is particularly true of ethnicity, where the data is missing or unknown for around a quarter of respondents to the survey; results split by ethnicity are not shown for non-UK domiciled students for this reason.

---

<sup>13</sup> For example, if male students are more likely to respond than female students, the survey responses will be skewed towards males.

<sup>14</sup> Available at <https://www.officeforstudents.org.uk/publications/update-to-associations-between-characteristics-of-students/>. This report was first published in 2019, and was updated in November 2020. We strongly value feedback as we continue to explore the best way to carry out and display analysis where multiple characteristics are considered. Any feedback on this report should be directed to [official.statistics@officeforstudents.org.uk](mailto:official.statistics@officeforstudents.org.uk).

## Raw data

19. The raw data that underpins the analysis is provided alongside the analysis itself. Both are available on the same webpage.<sup>15</sup>
20. The raw data is available in .xlsx format.
21. The fields included in the raw data are described in Tables 1 and 2. All decimal fields have been rounded to 2 decimal places.

**Table 1: Fields included in the raw data for ‘Sector results’**

Field	Description
Q	The number of the NSS question or scale.
QText	The text of the NSS question or scale.
Country	UK, England, Northern Ireland, Scotland, or Wales
Mode	All, Full-time or Part-time
Year	The year of the NSS survey.
Lower	The lower limit of the 95% confidence interval for the PerAgree field.
PerAgree	The proportion of respondents who selected ‘Definitely agree’ or ‘Mostly agree’.
Upper	The upper limit of the 95% confidence interval for the PerAgree field.
Resp	The number of students who responded to the question or scale.

**Table 2: Fields included in the raw data for ‘Results by characteristic’ and ‘Differences by characteristic’**

Field	Description	Calculated as
Q	The number of the NSS question or scale.	-
QText	The text of the NSS question or scale.	-
Type	The student or course characteristic corresponding to the row e.g. sex.	-
Description	The value of the student or course characteristic e.g. male, female.	-
Domicile	The value of domicile corresponding to the row i.e. UK or non-UK.	-
Year	The year of the NSS survey.	-
N_response	The number of students (FPE) who responded to the question or scale.	-

<sup>15</sup> See [www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/sector-analysis](http://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/sector-analysis).

Field	Description	Calculated as
Actual_prop	The proportion of respondents who selected 'Definitely agree' or 'Mostly agree'.	N_agree divided by N_response
Benchmark	The benchmark value for the value of the characteristic corresponding to the row.	See footnote <sup>16</sup>
Stdev	The standard deviation of the <i>Actual_prop</i> field.	See footnote <sup>17</sup>
Difference	The difference between the percentage of respondents who agreed and the benchmark.	Actual_prop – Benchmark
z	The number of standard deviations by which the percentage of respondents who agreed and the benchmark differ.	Difference divided by Stdev
Sig	The level of confidence we have that the difference between the agreement rate and the benchmark is significant. This value is also signed +/- to indicate whether the agreement rate is above or below the benchmark.	-

<sup>16</sup> See 'Technical notes': <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks>

<sup>17</sup> See 'Standard deviations' section: <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks>

**Table 3: Definitions of student characteristics and characteristics of higher education provision**

Characteristic	Categories <sup>18</sup>	Definition
Age	Under 21 21 – 24 25 and above	Age of student on the commencement date of their course
Sex	Male Female Other	Indicates the sex of the student
Ethnicity	Asian Black White Mixed Other	Indicates the ethnicity of the student
Disability status	Declared disability No known disability	Indicates whether the student has reported as disabled
Subject classification	Medicine and dentistry Subjects allied to medicine Biological and sport sciences Psychology Veterinary sciences Agriculture, food and related studies Physical sciences Mathematical sciences Engineering and technology Computing	Subject studied according to the Common Aggregation Hierarchy (CAH1) <sup>19</sup>

<sup>18</sup> 'Unknown' categories are not shown in this analysis.

<sup>19</sup> <https://www.hesa.ac.uk/innovation/hecos>



Characteristic	Categories <sup>18</sup>	Definition
	Architecture, building and planning Social sciences Law Business and management Language and area studies Historical, philosophical and religious studies Education and teaching Combined and general studies Media, journalism and communications Design, and creative and performing arts Geography, earth and environmental studies	
Mode of study	Full-time Part-time	Indicates the mode of study of the student

**Table 4: Confidence levels corrected for multiple comparisons for ‘Results by characteristic’ and ‘Differences by characteristic’**

Overall confidence level	Corrected confidence level for individual comparisons
0.5	0.9998299320
0.6	0.9998639456
0.7	0.9998979592
0.8	0.9999319728
0.9	0.9999659864
0.99	0.9999965986