

Methodology document for National Student Survey characteristic analysis

1. This document explains the methodology used to produce the National Student Survey (NSS) characteristic analysis published on 27 October 2020.

Background and scope

2. This data has been developed as an experimental statistic. It explores the extent to which responses to the NSS differ across student groups.
3. The data includes NSS results from the 2019 and 2020 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.¹ The analysis is at sector level, so no provider level results are included.
4. The data is presented in two interactive charts. These display the percentage of respondents who agreed (selected 'Definitely agree' or 'Mostly agree') to each question and scale² in the NSS. Responses have been separated according to six student and course characteristics: age, sex, ethnicity, disability status, 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.
6. A full description of the characteristics used in this analysis and their values is shown in Table 2. The data is shown separately for UK-domiciled and overseas students. Categories 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 27 October 2020.

Benchmarking and significance testing

7. We have used benchmarking to take into account factors that are known to influence NSS agreement rates. The benchmark is a weighted sector average. It gives information about the agreement rates that might be expected given the student and course characteristics included

¹ See pp. 7-8 of the NSS Good Practice Guide 2020, available at www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/promotion-of-the-nss/

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

in the weighting. Where differences exist between the rate and the benchmark, 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, disability status, subject classification and mode of study of respondents.³ 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.⁴
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 0.05. 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.

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.⁵ The Office for Students (OfS) is currently exploring different ways of showing statistical significance and we welcome feedback on the approaches taken here.⁶

³ 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.

⁴ 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.

⁵ 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.

⁶ Feedback should be directed to NSS@officeforstudents.org.uk.

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.⁷ To overcome this, an adjustment is often made to the calculation to control the false discovery rate (Benjamini and Yekutieli, 2001)⁸: in this case, a Bonferroni correction has been used.
12. We have considered the total number of comparisons as 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 overseas) 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.
13. A selection of corrected significance levels calculated using the Bonferroni method⁹ are shown in Table 3.

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 2020 results we examined the impact of the coronavirus (COVID-19) pandemic on the data.¹⁰ We did not find evidence that the results had been strongly impacted by the pandemic. While our analysis uncovered some potential issues with the results, such as the mode of response effect and the

⁷ 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.

⁸ 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

⁹ 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).

¹⁰ See www.officeforstudents.org.uk/publications/nss-2020-analysis-of-impact-of-coronavirus/.

variable response rates across student groups, these are all variations on themes present in previous years. We judge that they will not affect the reliability of this analysis.

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. Our approach to understanding and displaying significant differences is new for this analysis, and different from the approach used elsewhere in the NSS results. We have chosen not to include materiality as a factor when assessing significance. We have also not applied significance tests at a single fixed level; instead, the user can choose their desired level of confidence. We believe this is appropriate in this context because it allows users to decide the sort of differences they are interested in, rather than assuming that a particular sort of difference is relevant and interesting to all users. We have also used a different level of Bonferroni correction to take account of the large number of data items that can be considered in one viewing. We have been particularly careful to ensure the accuracy of these aspects of the analysis, due to their relative novelty.
19. 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 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. 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.¹¹ 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.
 - c. 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.
 - d. This analysis looks at students grouped by a single characteristic: age, sex, subject, mode, ethnicity or disability status. We acknowledge that this ignores the interplay between these

¹¹ For example, if male students are more likely to respond than female students, the survey responses will be skewed towards males.

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 'Associations between characteristics of students'.¹²

- e. Data about the characteristics of overseas students is not always available. This means that particular caution is needed in interpreting the results for overseas 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 overseas students for this reason.

Raw data

- 20. The raw data that underpins the analysis is provided alongside the analysis itself. Both are available on the same webpage.¹³
- 21. The raw data is available in .xlsx format.
- 22. The fields included in the raw data are described in Table 1. All decimal fields have been rounded to 2 decimal places.

Table 1: Fields included in the raw data

Field	Description	Calculated as
Q	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 overseas.	-
N_agree	The number of students (full-person equivalent (FPE)) who selected 'Definitely agree' or 'Mostly agree' to the question or scale. ¹⁴	-
N_response	The number of students (FPE) who responded to the question or scale. ¹⁴	-
Actual_prop	The proportion of respondents who selected 'Definitely agree' or 'Mostly agree'.	N_agree divided by N_response

¹² Available at www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students. This report was published in 2019, and a 2020 update will be published soon. 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.

¹³ See www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/sector-analysis

¹⁴ This will be a whole number for all characteristics except subject of study, where it may be a decimal. This reflects the fact that some students' study activity is split across multiple subject areas. This may also be a decimal for scales.

Field	Description	Calculated as
Benchmark	The benchmark value for the value of the characteristic corresponding to the row.	See footnote ¹⁵
Stdev	The standard deviation of the <i>Actual_prop</i> field.	See footnote ¹⁶
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
Year	The year of the NSS survey.	-
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.	-

¹⁵ See 'Technical notes': <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks>

¹⁶ See 'Standard deviations' section: <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/benchmarks>

Table 2: Definitions of student characteristics and characteristics of higher education provision

Characteristic	Categories ¹⁷	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 Architecture, building and planning Social sciences Law Business and management Language and area studies Historical, philosophical and religious studies Education and teaching	Subject studied according to the Common Aggregation Hierarchy (CAH1) ¹⁸

¹⁷ 'Unknown' categories are not shown in this analysis.

¹⁸ <https://www.hesa.ac.uk/innovation/hecos>

Characteristic	Categories ¹⁷	Definition
	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 3: Confidence levels corrected for multiple comparisons

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