National Student Survey

Assessment of alternative sample designs

Enquiries to nssreview@officeforstudents.org.uk

Publication date 30 March 2021

Contents

Introduction

1. As part of the 2020 review of the National Student Survey (NSS),[1] we assessed whether we could collect student experience data without conducting an annual census survey. We considered six alternative survey designs – an overview of our assessment is shown in Table 1.

Table 1: Assessment of different survey designs

		Financial costs	Compliance costs for providers	Respondent burden	Data needs of users		
					Students	Providers	Funders/ regulators
A	Small stratified sample						
B	Large stratified sample						
C	Biennial census						
D	Undergraduate sample						
E	Volunteer survey						
F	Opinion poll						

Key

Much better than the present design
Slightly better than the present design
Roughly the same as the present design
Slightly worse than the present design
Much worse than the present NSS design
The evaluation is highly tentative

2. This document provides more details about each sample design, and explains how we arrived at our assessment. Throughout, we consider whether the alternative samples would meet the needs of users of the NSS data. These needs were identified through the wider NSS review, and are described in Annex A.

3. Further information is available from Will Hanson (Will.Hanson@officeforstudents.org.uk) and Anna Sherratt (Anna.Sherratt@officeforstudents.org.uk).

[1] The main NSS review is available at: www.officeforstudents.org.uk/publications/nss-review-phase-one-report/.

Background

Notes on method

4. Throughout, we consider alternative ways of sampling **full-time** undergraduate students. We have not extended our evaluation to the part-time student population. This is simply because full-time students present the easier challenge. There are more of them, and as a result it is easier to arrive at a sample design which supports reliable estimates without surveying the whole population. If we decide to proceed with further exploration of any of the options, we will need to consider whether or not the method will work for part-time students, or whether an alternative approach is needed (for example, it would be possible to run a stratified sample for full-time students alongside a full census survey for part-time students).

5. The evaluation of the options is against the current, annual NSS census. To make this evaluation, we assume that everything else remains unchanged. For example, we assume that the questionnaire has the same length and format as the current NSS. And we assume that the target population remains the same – unless changing this population is specified as part of the design (as for design D). These are only working assumptions, made to allow us to understand the advantages and disadvantages of each survey design without distraction. In reality, many other elements of the survey will be considered as part of the NSS review, and these decisions will also affect the costs and burdens of the survey, and the accuracy of the data it provides.

Note on compliance costs and 'indirect costs'

6. As part of our evaluation, we considered whether the six survey designs would reduce compliance costs for providers. By compliance costs, we mean the things that providers need to do in order to participate in the survey, such as supplying contact details or encouraging students to respond. We have not attempted to quantify compliance costs exactly, but instead have noted whether each survey design reduces or increases the requirements placed on providers.

7. Many providers also engage in NSS-related activities which are not formally required by participation in the survey. These include carrying out additional analysis to address their own areas of interest, and implementing strategies to improve NSS results. Some of these activities are controversial. In our roundtable discussions, some academic staff told us that the NSS encourages university management to target 'quick wins', and distracts academics from more substantial improvements in teaching. In contrast, university managers and administrators tend to see the NSS as a helpful tool for improving the student experience.[2]

8. A standard test of whether A causes B is to consider whether B would occur in the absence of A. Our research found that many university staff believe that if the NSS did not take place,

[2] As part of the 2020 NSS review, we surveyed staff at UK higher education providers. 68 per cent of respondents who described themselves as working in administrative, support or management roles agreed that the NSS has contributed to improving the student experience, compared with 33 per cent of respondents who described themselves as working in academic roles. In total, 990 respondents gave a substantive response to this question. This finding was borne out in our roundtable discussions with providers.

providers would create their own bespoke surveys.[3] They do this because they see understanding the student perspective as a key element of improving learning and teaching. This suggests that the controversial activities would continue even if the NSS did not exist, or if it provided less data. The activities are not caused by the NSS itself, but by an approach to managing quality. The approach is presently enabled by the NSS data, but could equally be supported by in-house data (which would itself be burdensome to collect).

9. In the evaluation of alternative survey designs, we do not take into account 'indirect costs': the time that providers spend on NSS-related activities that are not formally required or expected. We see these as resulting from decisions by university leaders and managers, rather than caused by the survey. Nonetheless, it is worth noting that some designs enable these activities in a way in which other designs do not. In particular, designs A, E and F do not support estimates about the student experience at a course level. The data they produce therefore cannot be used to make changes to particular courses or departments. This does not mean that university managers would no longer use student experience data to assess and influence teaching; simply that the data would need to be sourced elsewhere.

[3] In our survey of staff at UK higher education providers, 93 per cent of 900 respondents said that if the NSS did not take place, it is likely that their institution would replace it with their own survey. This was corroborated by our roundtable discussions, and also by what we know about current behaviour. Many providers run surveys to collect data about courses not covered by the NSS, including sometimes replicating the NSS questions.

Design A: Small stratified sample

Design

10. The full-time eligible survey population is stratified into providers, and then again further stratified by CAH1 subject groups.[4] Students are randomly selected from these sub-strata. Sampling is initially proportionate – that is, all students have an equal chance of selection – but subject to a provider floor. The provider floor ensures that at least 50 students (or the whole population) are selected at each provider. In total, around 30,000 students are selected. The probability of selection is between 6 per cent and 100 per cent. Design weights are applied when calculating estimates from the survey results, to compensate for the over-sampling of some students due to the provider floor.

11. Provider and subject of study both have a large impact on the student experience. By including these factors as strata, we reduce the risk of uncontrolled variance – that is, misleading results due to an unbalanced mix of students in our sample selection. The provider floor is included because it ensures that we sample enough students at each provider to be able to draw conclusions about the student experience at a provider level.

12. This sample design was investigated as a way of exploring the potential of a sample that was relatively small, but still large enough to generate estimates for every provider (something that has been identified as a priority by data users).

Costs and burden

13. We estimate that design A would cost about half as much to implement as the current survey design. This estimate is arrived at by keeping fixed the current costs of the survey, and adjusting the costs that vary with the number of respondents. It also takes into account advice from the current survey contractor, to the effect that the current survey takes advantage of efficiencies of scale which would be lost with a smaller survey.

14. Design A would require providers to check target lists and supply contact details, although for a smaller number of students: this would slightly reduce the workload. With such a small proportion of students sampled, it would not be appropriate to require providers to promote the survey. Understanding the results would be harder, as providers would need to take into account the uncertainty introduced by the sample design. Some providers would feel the need to understand and question the sampling method, which would add to burden. On the whole, we judge that design A would lead to a slight reduction in compliance costs for providers.

15. Respondent burden is measured as the response time per respondent multiplied by the number of respondents.[5] When comparing this design with the current survey, we will assume that the response time per respondent remains constant, as does the proportion of non-

[4] CAH1 is the highest level subject grouping. See https://www.hesa.ac.uk/innovation/hecos for more information.

[5] We used the Government Statistical Service guidance 'Monitoring and reducing respondent burden', published on 10 November 2020. Available at https://gss.civilservice.gov.uk/policy-store/monitoring-and-reducing-respondent-burden-2/.

respondents. The reduction in response burden therefore directly reflects the reduction in sample size: that is, it is around seven per cent of that of the current survey.

Accuracy

16. We assessed the sampling error introduced by the survey design. The sampling error is the variation in the results which occurs because we have sampled a subset of students, rather than surveying the whole population. To assess this, we considered what would have happened if, in 2019, we had implemented the NSS using the design A. That is, had the survey been based on a stratified sample of 30,000 students, how would the results have differed from those we actually collected?

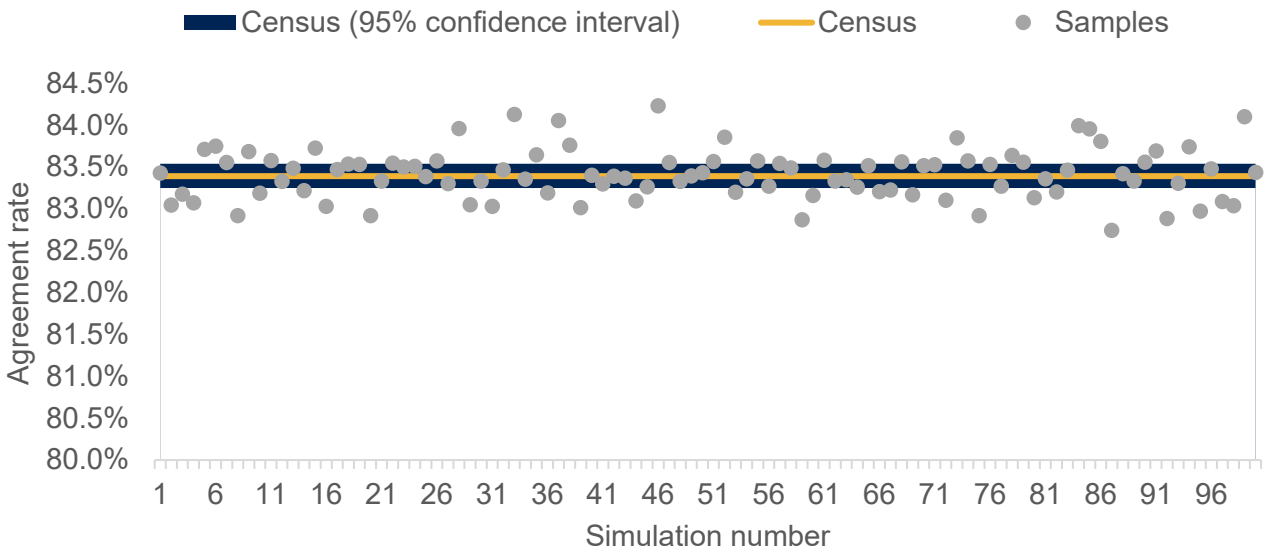
17. Since the survey design involves random selection, it creates a different sample each time it is applied. To allow for this, we used the design to create 100 different samples from the 2019 target population. Since each eligible student was surveyed as part of the 2019 survey, each sample has a corresponding set of survey responses (and non-responses). We used these responses to generate key statistics from each sample, such as the agreement rate for question 27 (overall satisfaction). We compared these with the results from the full census survey to understand the sampling error introduced by the design.

18. Figure 1 shows the comparison for Question 27[6] (overall satisfaction) for England. The agreement rate generated by the full census survey was 83.4 per cent: this is shown by the yellow line. The blue area running across the chart show the 95 per cent confidence interval around this statistic. This is the area within which we are confident that the true value falls. Each dot on the chart represents one of the estimates of overall satisfaction generated the samples. Figure 1 shows that if design A had been used to run the 2019 NSS, the likely range of estimates would span 82.9 per cent to 84.0 per cent.[7] Sixty of the 100 estimates are outside the 2019 confidence interval: that is, they are not within the range of values that we believe to contain the true value of student satisfaction.

[6] Question 27: Overall, I am satisfied with the quality of the course.
Response scale: Definitely agree; Mostly agree; Neither agree nor disagree; Mostly disagree; Definitely disagree; Not applicable

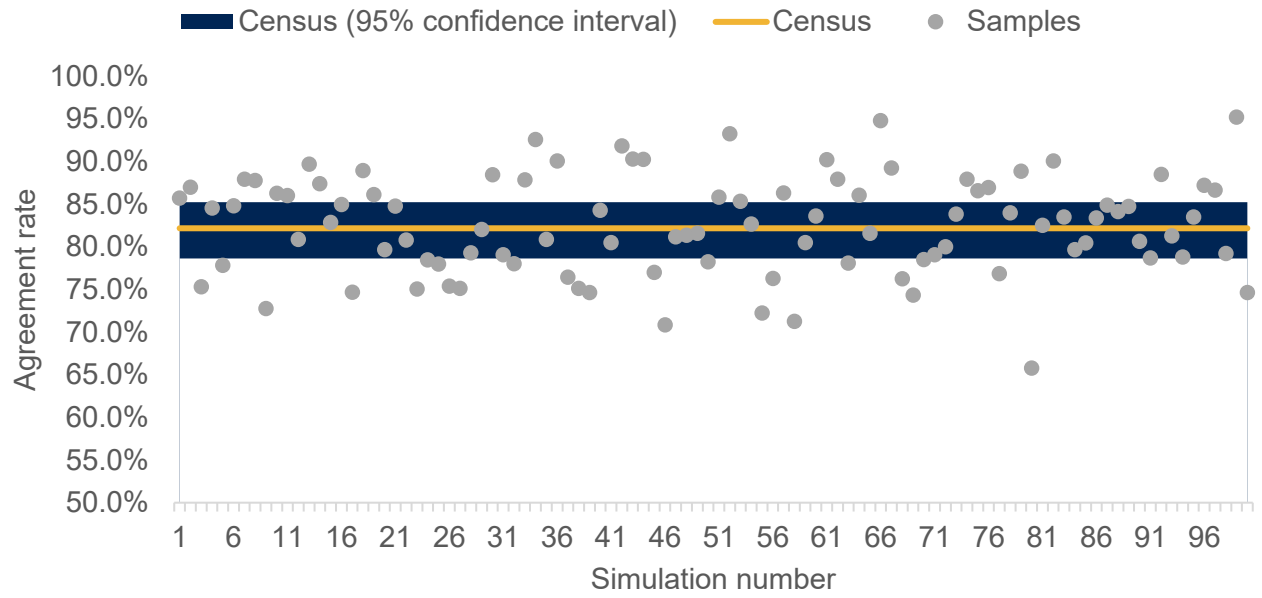
[7] Here and elsewhere in this document, we use the fifth to ninety-fifth percentile range to give the likely range of estimates. Roughly speaking, this means that nine times out of ten, the survey design would produce an estimate within this range. The rest of the time the estimate is outside this range.

Figure 1: Question 27 estimates for England (design A)



19. We repeated this experiment, this time looking at results for individual providers. Since more than 400 providers participate in the NSS, 100 samples per provider generates a lot of data. In Figure 2, we show the results of the experiment for a single, fairly large provider (with around 600 students eligible to respond to the survey). As above, the horizontal lines show the confidence interval around the estimate from the 2019 NSS, and the dots show the estimates generated from each sample. In only 39 cases do the estimates fall within the confidence interval of the full census result. The likely range for the estimates is 74.3 per cent to 90.4 per cent compared with a full census value of 82.2 per cent.

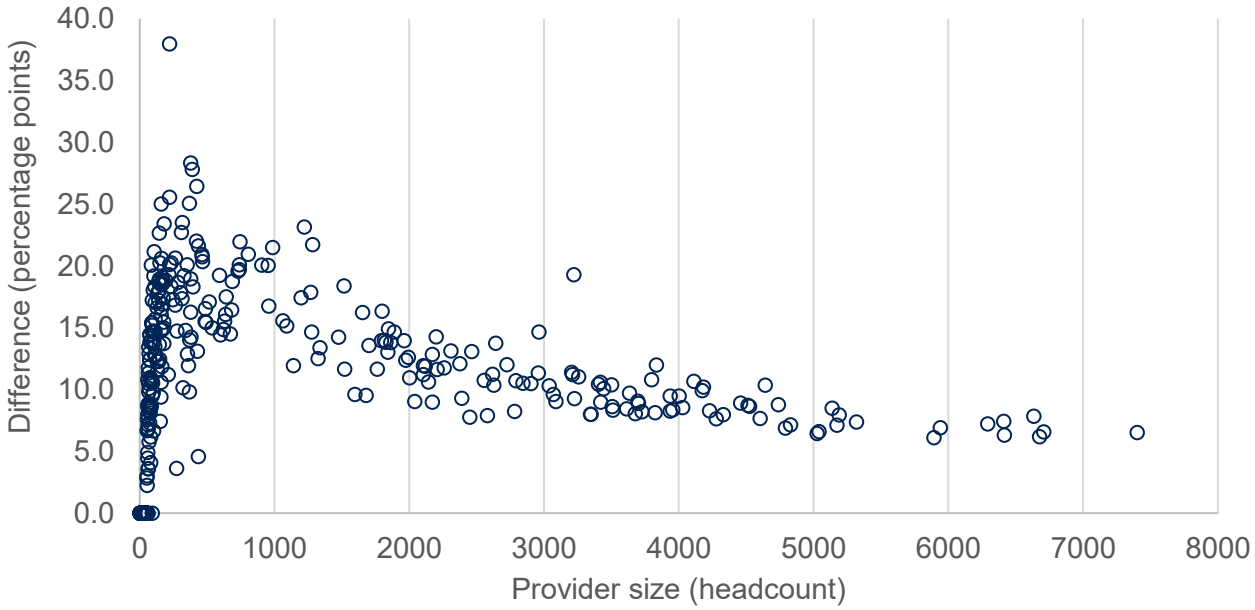
Figure 2: Question 27 estimates for a single provider (design A)



20. Figure 3 shows similar data, but this time including all providers. The value shown for each provider is the likely range of the estimates generated by the sample design – that is, the difference between the fifth percentile and the ninety-fifth percentile. For the very smallest

providers (at the far left-hand side of the chart), the range is zero. This is because all, or nearly all, of their students will have been selected in the sample, leaving no room for sampling error. For larger providers, but with fewer than 1000 eligible students, the range is large: typically, between 10 and 25 percentage points. For the largest providers, the range decreases again although remaining at around six percentage points. This is because the larger number of students selected from the larger providers makes their estimates more stable (even though the probability of selection remains roughly the same).

Figure 3: Likely range for Question 27 estimates for all providers (design A)



Availability of data

21. We also tested whether the survey design provided enough data to support conclusions about the student experience at course level. We use the term 'course level' loosely, to refer to any data about a subject within a provider, even if it is not specific to a particular course. Here we have simply tested for the availability of data, rather than assessing sampling error. We take data to be available if the sample generates responses from 10 or more students on the course.[8]

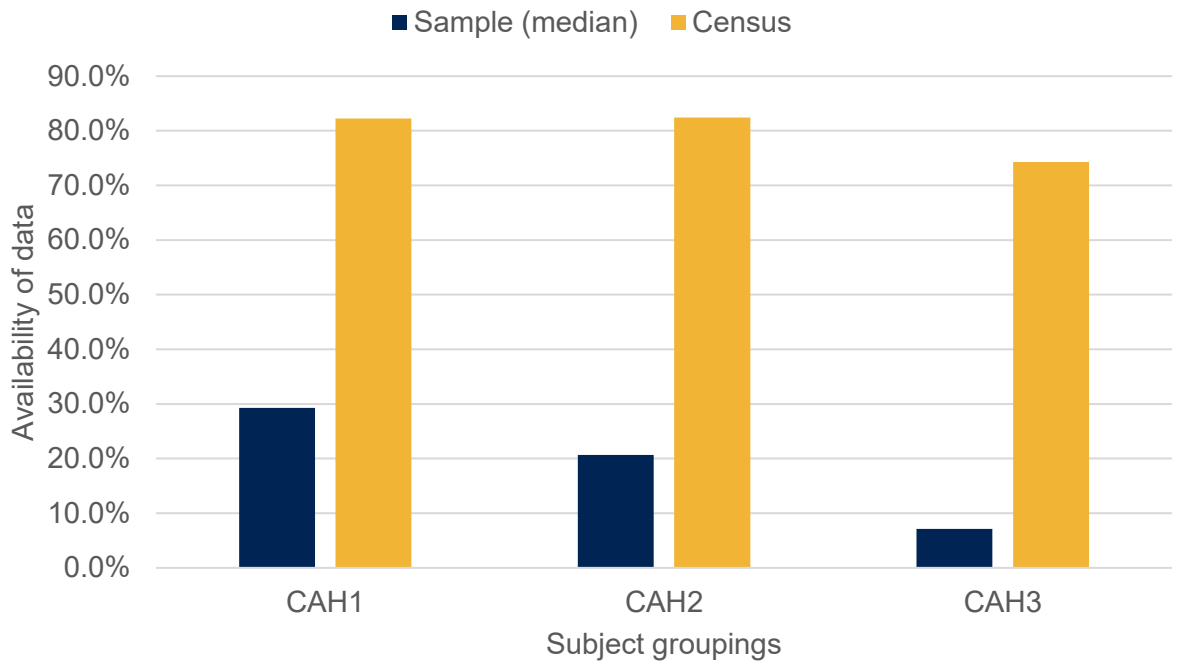
22. Figure 4 shows the proportion of courses with data. The subject groupings are those of the common aggregation hierarchy, which become more specific as we move from left to right. For example, CAH1 includes Engineering and Technology, CAH2 includes Engineering, and CAH3 includes Mechanical Engineering.[9] The yellow bars show the percentage of courses with data

[8] There is a general challenge here. Data users tell us that they want very granular data, which often entails estimates based on small populations. From a statistical perspective, this data will involve a high degree of uncertainty – that is, there is a risk that the estimates it produces will differ considerably from the truth about the population of interest. Since this challenge is present to some extent with the current NSS as well as the alternative survey designs, we have set it to one side for the purpose of this evaluation. We remain committed to finding ways to control and communicate the uncertainty related to small numbers.

[9] For simplicity, students studying in two different CAH subject groups have been excluded from the calculation. This applies to proportions from both Design A and the census survey. We judge that if these students were included both proportions would rise slightly, but that the contrast between them would remain.

for the full census NSS. The yellow bars show a comparable figure for design A. Since we have run design A 100 times, the proportion for design A is the median of the 100 proportions generated by these runs. From Figure 4, we can see that design A would supply CAH3 data for around seven per cent of courses, compared with 74 per cent from the full census NSS.

Figure 4: Availability of data about courses (design A), compared with census survey



Design A conclusions

23. Design A roughly halves the direct cost of the survey, and brings smaller reductions in compliance costs for providers.

24. The sample design might be used to understand the student experience in England, provided that we accept that small changes (of 1 or 2 percentage points) could not be detected with confidence. However, it does not provide useable provider-level estimates. As Figures 2 and 3 show, we can get very different estimates for a single provider simply because of the students selected for the sample. This means that we cannot be confident in detecting large differences. For example, for the provider shown in Figure 2, we could not be confident that a drop from 90 per cent to 73 per cent was a real change, as we know that this variation could be purely due to sampling error. (Note that this presentation of the issue ignores other sources of uncertainty, which will make it harder still to detect differences with confidence.)

25. Neither does the sample design support course-level data (even when this is understood very loosely as information about high-level subject groupings within a provider). There is simply not enough data collected for this, even setting aside questions about accuracy.

26. While design A, or something like it, could be used to understand the student experience in general – across the UK, or across England – it does not supply the more detailed information that is needed by most data users. We therefore recommend that this design is not taken forward.

Design B: Large stratified sample

Design

27. Design B is a larger version of design A. The eligible population is stratified by provider and then by CAH1 subject group. Random sampling is used to select students from the sub-strata. Sampling is generally proportional, but subject to a provider floor of 50. The overall sample size is 250,000. The probability of selection is between 60 per cent and 100 per cent.

28. We investigated design B because we found, through experimentation, that it is the smallest sample that provides provider-level estimates with acceptably low levels of sampling error – that is, low enough that the estimates from the sample would meet the needs of most data users.

Cost and burden

29. We estimate that the cost of design B would be about 75 per cent of the current cost. As for design A, this estimate is based on adjusting the costs of the current survey to take into account the lower student numbers, while recognising that some fixed costs remain static.

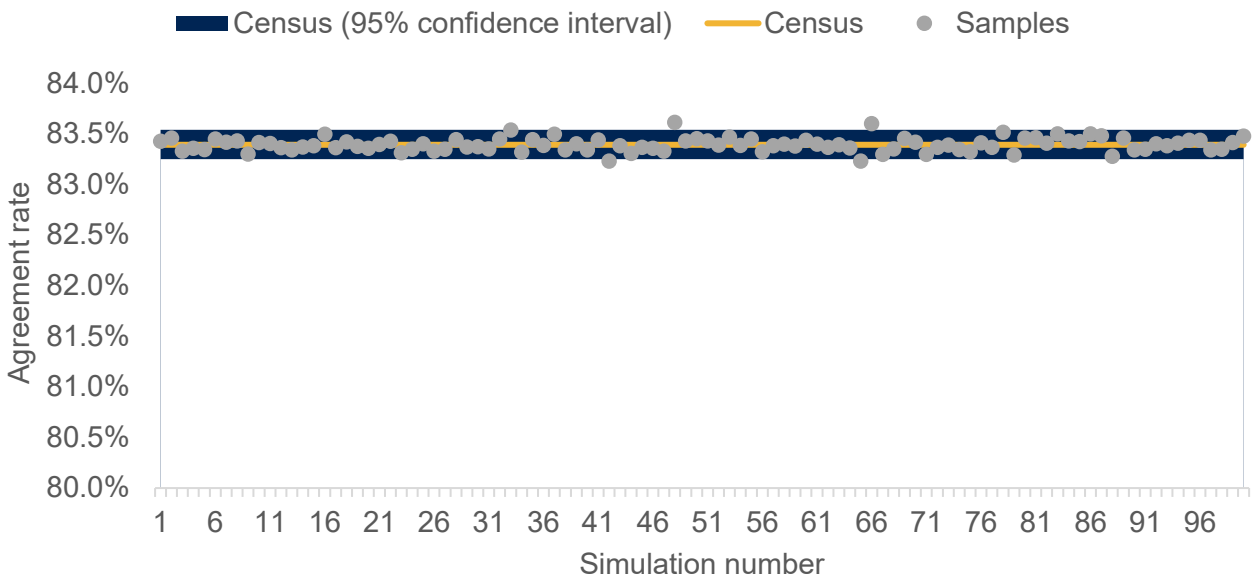
30. Like design A, this design would require providers to check target lists and supply contact details, although for a reduced number of students compared with the full census. Since a large proportion of students would be surveyed at each provider, it might be appropriate to ask providers to promote the survey; or alternatively, all promotion could be done centrally to reduce burden. As with design A, understanding the results would be more complex, and some providers would wish to interrogate the sampling method. On the whole, we judge that design B would lead to a slight reduction in compliance costs for providers.

31. The respondent burden would be around 60 per cent that of the present survey.

Accuracy

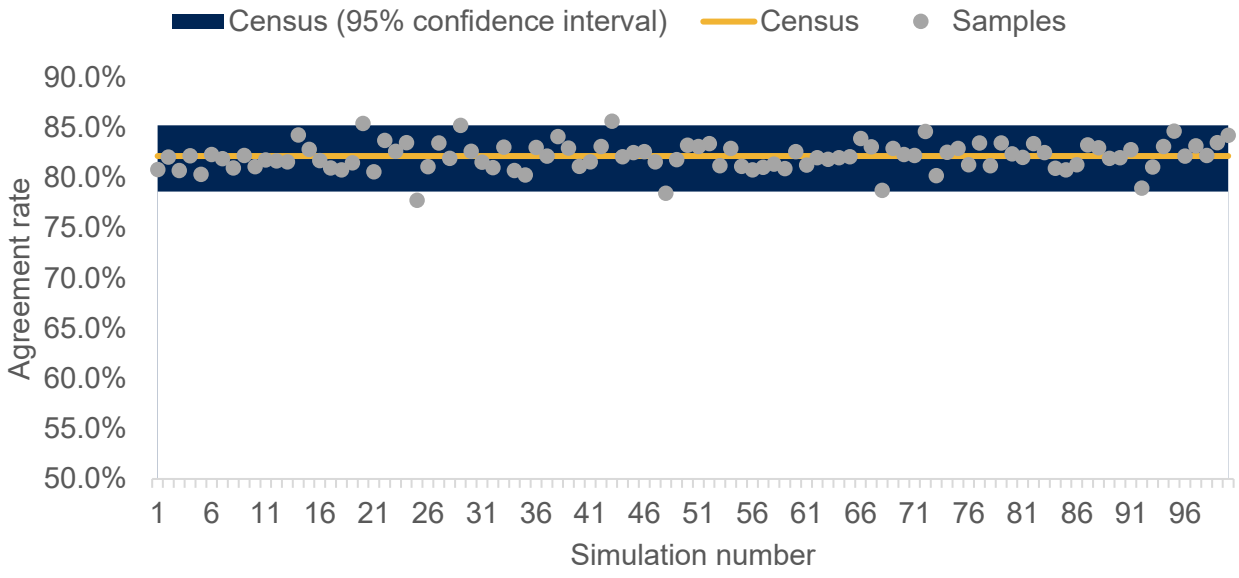
32. We assessed the sampling error introduced by the survey design, using the approach described in paragraphs 16 and 17 above. Figure 5 shows the multiple estimates of the Question 27 agreement rate for England derived from 100 applications of the sample design, compared with the estimate from the full census survey. We can see from this, that compared with design A, design B yields estimates that are very close to the results from the full census survey: that is, the sampling error is small. The likely range of estimates falls within the 95 per cent confidence interval of the full survey result.

Figure 5: Question 27 agreement rate estimates for England (design B)



33. Figure 6 shows the estimates of the Question 27 agreement for a single provider (the same provider is used as an example throughout to maintain consistency). Again, the design B estimates are closer to the full census result. Ninety-eight of the 100 estimates fall within the 95 per cent confidence interval. The likely range of the design B estimates is 80.2 per cent to 84.3 per cent.

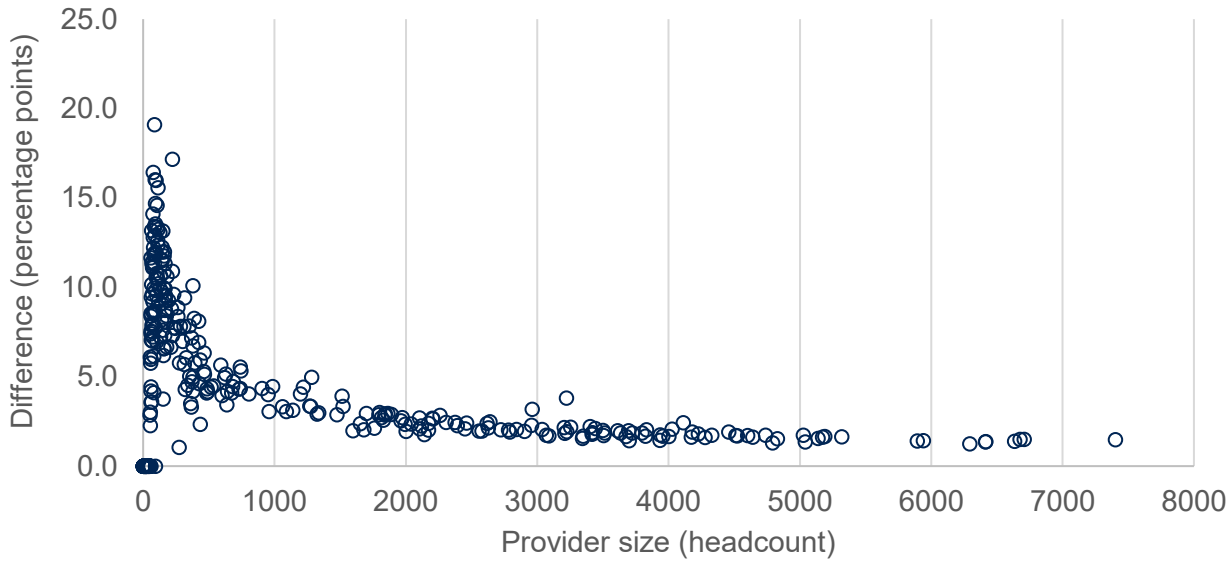
Figure 6: Question 27 estimates for a single provider (design B)



34. Figure 7 generalises this to show the likely range of the same estimate for all providers. For smaller providers, the ranges are very small (as they are with design A): this is because the provider floor ensures that all, or almost all, eligible students are included in the sample. As we move from smaller to larger providers, the ranges increases as the probability of selection reduces, and then drops again as the probability of selection stabilises and the number of

students selected grows. For most providers (more than 80 per cent) the likely range of estimates is smaller than 10 percentage points.

Figure 7: Likely range for Question 27 estimates for all providers (design B)

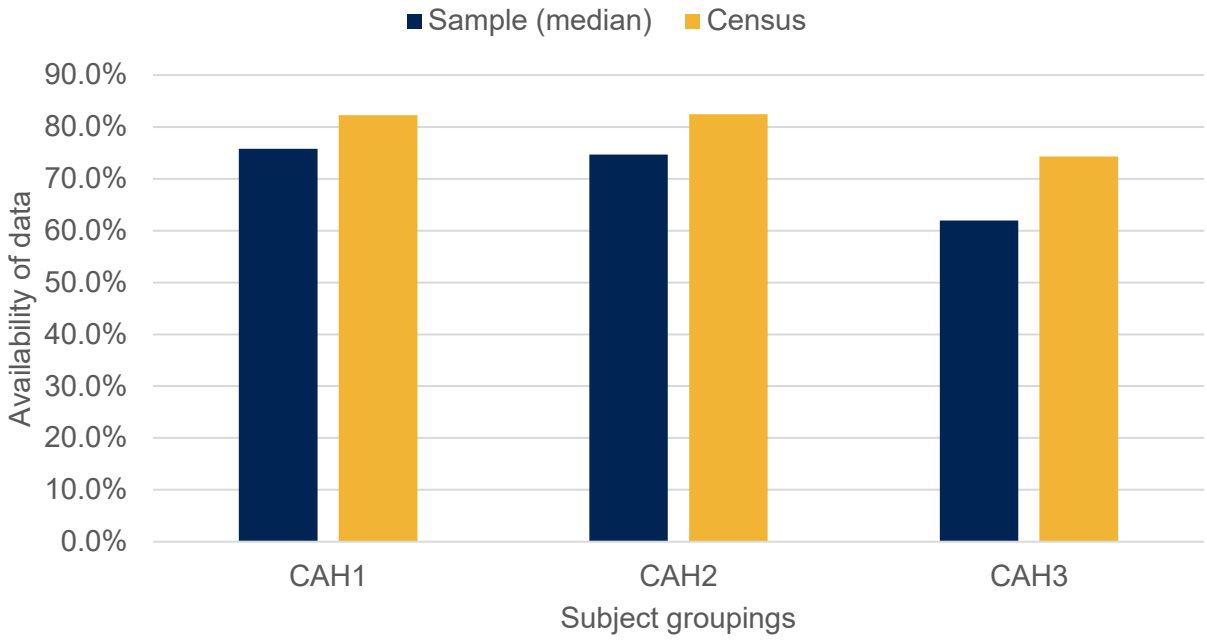


Availability of data

35. Figure 8 shows the availability of data about courses, as described in paragraphs 21 and 22. The sample design provides more course-level data than design A, but still falls slightly short of the standard reached by the full census survey. For example, design B provides information about 61.9 per cent of CAH3 units within providers, compared with 74.3 per cent from the full census survey.[10]

[10] As noted in paragraph 21 and the accompanying footnote, we have not yet considered the accuracy of the data of the data about courses, as opposed to the availability. A further step would be to consider the extent to which the sample design compounds the uncertainty already present due to issues around small populations.

Figure 8: Availability of data about courses (design B), compared with census survey



Design B conclusions

36. Design B comes far closer to meeting the needs of data users than design A. Due to the much larger sample size, it is possible to make fairly accurate estimates about providers. For most providers, the risk of sampling error would not prevent us from detecting large changes in the student experience.

37. Design B provides information about courses, although this is strongest when we consider the highest level subject groupings. The proportions drop when we look at the most detailed subject groupings; and there will be less data still available about specific courses. Nonetheless, this design would provide at least some of the data that students and prospective students want to see, and that staff at providers find useful.

38. Design B might be a workable alternative to a full census survey, although further work is needed to understand the extent to which the design yields accurate estimates for other groups of interest: equality groups, for example. The design would be more complicated than the census survey to implement; and the estimates would be harder to understand (and probably more contested). In deciding whether to implement this design, we should consider whether the extra complexity and risk are worth the benefits, particularly given that the design yields only a small reduction in compliance costs for providers.

Design C: Biennial census survey

Design

39. Design C is a full census biennial survey of final year students. At its simplest, all providers could be surveyed every two years. As a variation on this, half of providers could be surveyed each year. An advantage of the variation is that, with careful selection of the provider groups, the survey would support yearly national estimates. A disadvantage of the variation is that it would compromise the comparability of providers, particularly when one year was affected by an extraordinary event (such as a pandemic).

40. Our assessment focuses on the simple version of the biennial survey (all providers surveyed every two years), although many of the points made apply to the variation too.

Costs and burden

41. By running the survey biennially, financial costs would be approximately halved.

42. Providers would need to do everything that they do for the current national student survey, but only once every two years. Again, this approximately halves compliance costs, although there might be some inefficiencies related to staff turnover, and loss of knowledge in the gap between one survey cycle and the next.

43. Average annual response burden would also be halved.

Data availability and accuracy

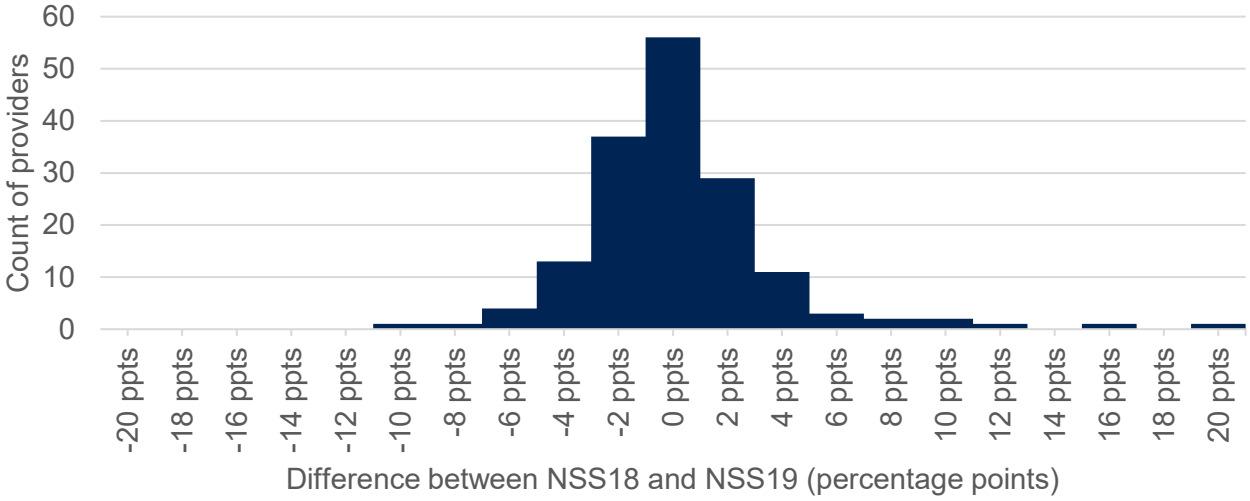
44. Design C raises very different issues from the other survey designs. The design is not prone to sampling error. Since we are surveying all eligible students, our conclusions cannot be distorted because of the sample selection. However, we face a new quality risk related to time lag. The NSS results are not only used to provide information about the student experience in a given year. They are also used to draw conclusions about the provider more generally, and about what the student experience will be like at the provider in the future. When a prospective student uses the NSS results to inform a decision about which course to apply to, they are (usually implicitly) using information about the past to infer something about the future: about what their student experience might be like if they studied at the provider.

45. At present, the NSS statistics already have a time lag. Students fill in the survey between January and May; the results are published in July; they remain current until the following July. Moreover, students are encouraged to reflect on all years of their course. For a student on a three-year course, the results published in July reflect a student experience that started nearly three years earlier. A biennial NSS would increase this time lag – at the end of the survey cycle, the estimates would use survey responses submitted more than two years previously.

46. This will make a difference. NSS results are mainly stable year on year, but it is not unusual to see large changes, both at provider level and at course level. Figure 9 shows the change in the Question 27 agreement rate at provider level between 2018 and 2019. While most providers saw no change or very little change, a small handful of providers saw changes of more than 6 percentage points (either up or down). For these providers, viewing their 2018 estimates rather

than their 2019 estimates would lead us to draw very different conclusions about the student experience.

Figure 9: Provider differences in Question 27 agreement rate, NSS19 compared to NSS18



47. To fully assess the impact of the time lag, we need to do more work to understand the stability of the data, and the expectations of users. For example, insofar as the data is used to draw conclusions about the future, it is already subject to a time lag problem: does adding an extra year to the time lag make a material difference? When the data is used by regulators and funders, is it important that the data is as up-to-date as possible, or could a two-year time lag be tolerated? We have recommended that these questions are explored further as part of feasibility work on design C.

Design C conclusions

48. A biennial survey is a simple way to reduce the costs and burden of the NSS. It avoids sampling error. In this respect, the results will be more powerful, and easier to understand, than those produced from a sample survey. However, the time lag introduces new problems. We propose further feasibility work to understand these, and their impact on data users.

Design D: All undergraduates

Design

49. Design D extends the eligible population of the NSS to all undergraduate students, but uses a sampling approach rather than a full census. We explored this design because, during the NSS review consultations, many data users expressed an interest in information about a wider population of students. Our sample design uses providers as strata, year of study as sub-strata and CAH1 subject groupings as a further level of sub-strata. For simplicity, our years of study are: final year, first year, all other years. This is an over-simplification which would need to be improved if we implemented this design.

50. Through our work on design B, we established that a sample of around 250,000 would be required to gain the information we need about final year students. Extrapolating from these, we would need a sample of roughly 750,000 students to gain similar information about all three of our year groups. This is not an acceptable proposal in the context of work to reduce the burden and the costs of the NSS. Instead, we approached the problem from a different angle: we asked, what could be achieved by keeping the sample size roughly the same but spreading these numbers over all years of undergraduate study? This approach results in around 150,000 students surveyed in each year group: 450,000 students in total. The probability of selection ranges from 30 per cent to 100 per cent.

51. In designing the sample, we have assumed that all estimates from design D will be split by year group, rather than aggregated together to form a single statistic. This is because, without further research, we do not know whether students from different years groups would interpret survey questions in the same way; or even whether they should be asked the same questions. We therefore cannot assume that it will be possible to provide a single estimate for all undergraduates. Testing and piloting might show us that it is possible to design questions that apply well, and consistently, to all year groups. If this is the case, then we might be able to design a smaller undergraduate sample which provides reasonably reliable data (although at a cost of less accurate information about how the student experience varies across years of study).

Costs and burden

52. We estimate that design D would cost a little more to run than the current NSS. The extra cost would be due to the innovations: for example, implementing the sampling method and designing questions and promotional activity that would apply to all year groups.

53. The compliance costs for providers would also be slightly higher, in particular due to the extra work required to provide information about first year students. (NSS target lists are generated from administrative data submitted by providers to HESA and the ESFA. This data is not submitted until after the end of the academic year, meaning that at the point when we run the survey, we have no information at all about students who have recently started a course. HESA's Data Futures project is working to develop in-year data streams, which would remove this additional complication.)

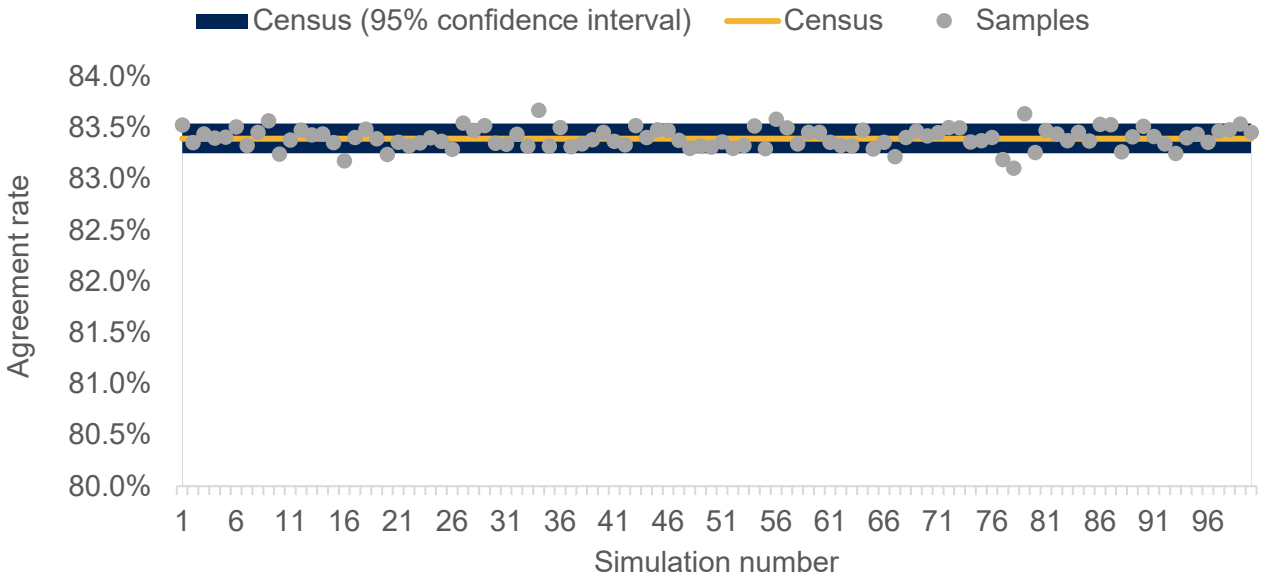
54. The respondent burden associated with design D would be roughly the same as for the current survey design.

Accuracy

55. We cannot explore the design D sampling error in the way in which we did for designs A and B. We have never collected large scale student experience data from non-final year students, and so we cannot model what would happen if we had implemented the sample design in previous years. Instead, to understand the sampling error associated with design D, we have focused on the results for the 150,000 final year students which would be included in the sample. In the absence of other data, our best guess is that the sampling error for the other year groups would be of a similar order.

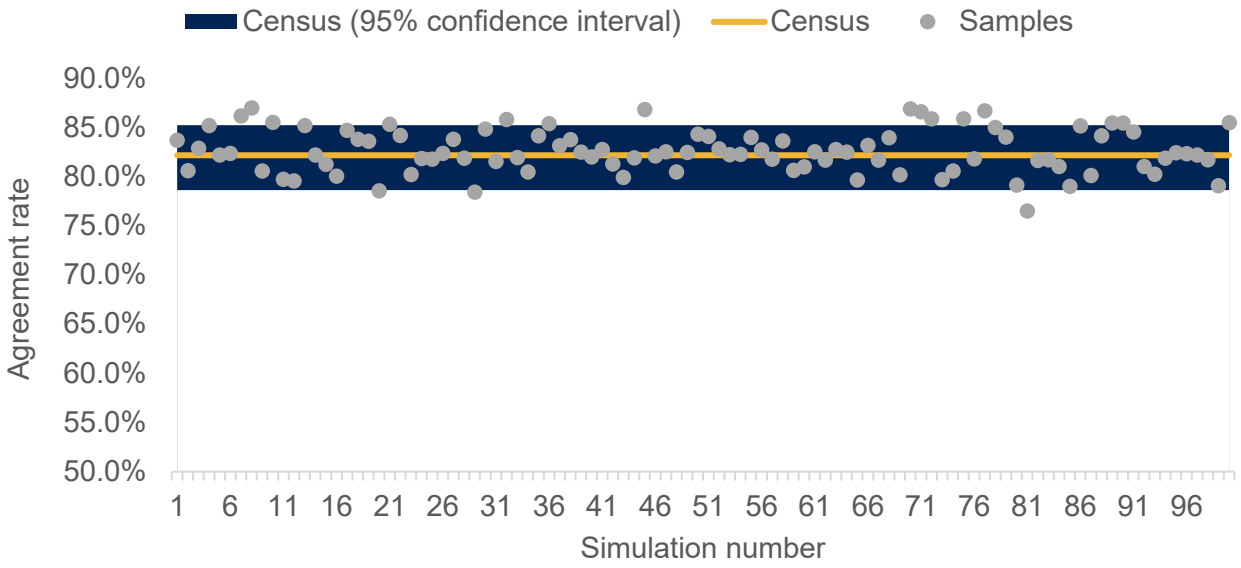
56. Figure 10 shows the estimated Question 27 agreement rate for final year students in England derived from 100 samples created using design D. The likely range of results is between 83.4 per cent and 83.5 per cent, with 12 estimates falling outside the confidence interval of the result from the full census survey.

Figure 10: Question 27 agreement rate estimates for England (design D)



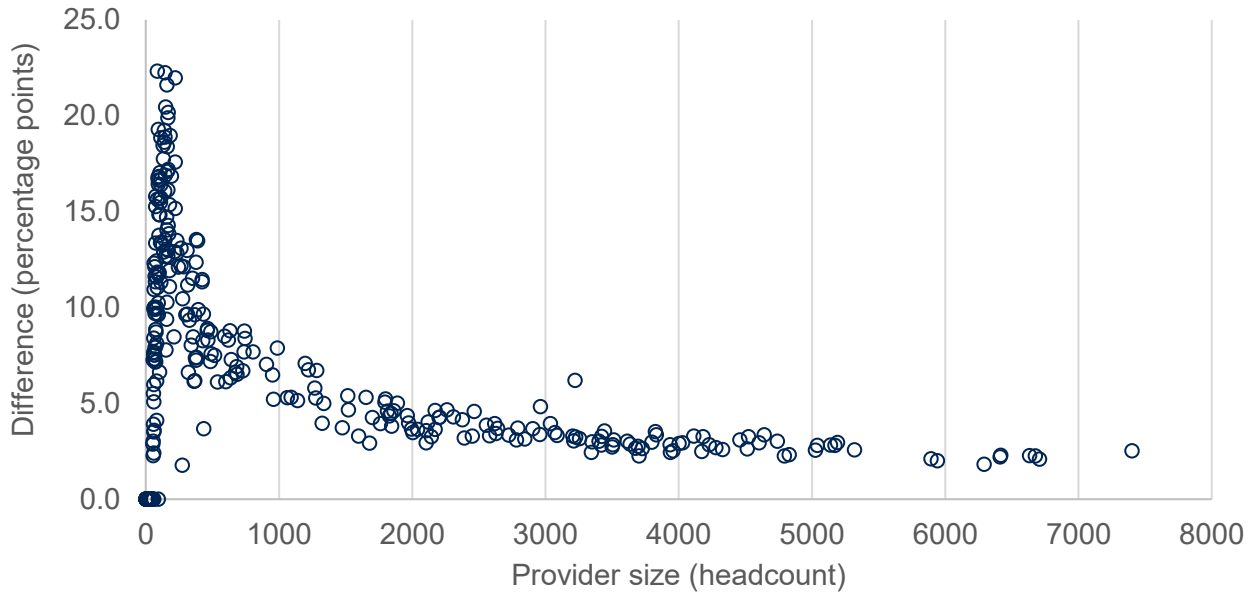
57. Figure 11 shows the estimated Question 27 agreement rate for final year students at a single provider. The likely range of results is between 79.1 per cent and 86.2 per cent, with 18 estimates falling outside the confidence interval of the full census survey.

Figure 11: Question 27 estimates for a single provider (design D)



58. Figure 12 shows the likely range of results for all providers. For the smaller providers (but with more than 50 eligible students), the range can be as great as 22 percentage points. For most larger providers with more than 1000 eligible students, it narrows to five percentage points or lower.

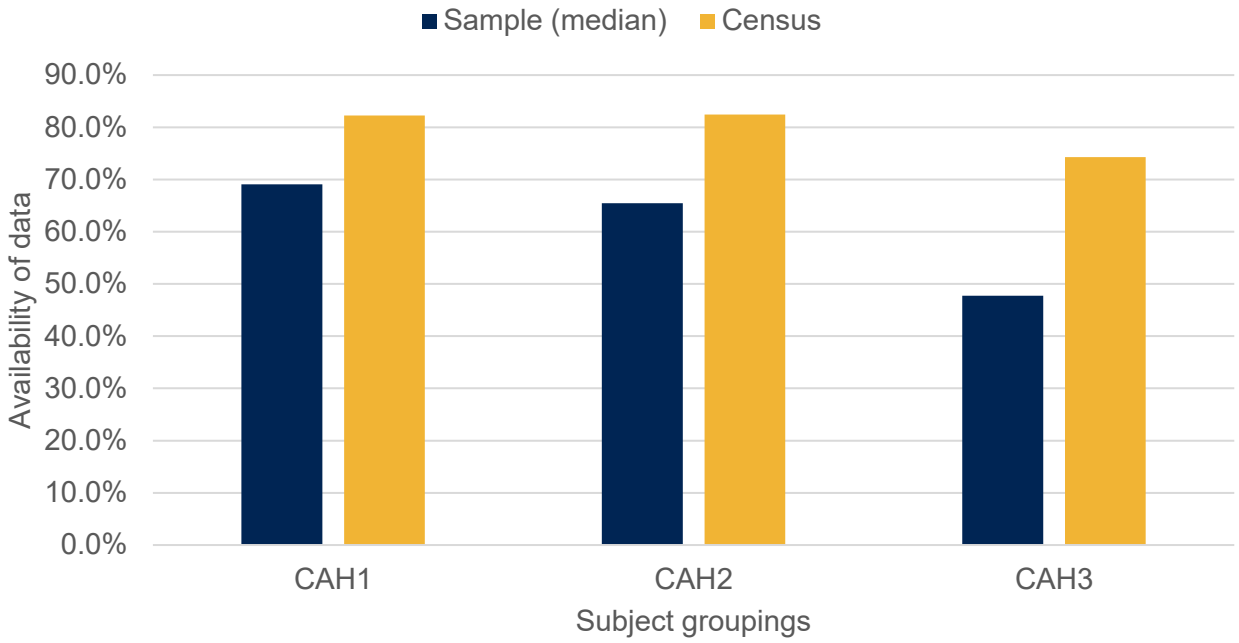
Figure 12: Likely range for Question 27 estimates for all providers (design D)



Availability of data

59. We assessed the availability of data about courses for the final year students surveyed as part of design D. This is shown in Figure 13. Looking at CAH1 data within providers, around 69 per cent of units meet the data availability threshold. When we focus on more detailed data which relates more closely to specific courses, the availability drops to 49 per cent.

Figure 13: Availability of data about courses (design D), compared with census survey



Design D conclusions

60. Our findings for design D are tentative. However, generalising from what we see when modelling the survey for final year students, design D would provide acceptably accurate data at national level for use by regulators, funders and other interested bodies. As with our other designs we see there can be substantial variation in the figures generated for smaller providers above the provider floor, while estimates for larger providers are more reliable.

61. When looking at the availability of data relating to courses, we see that design D provides substantially less data at subject level within providers when compared to design B and the census. It provides figures for more than half of provider subject groupings at common aggregation hierarchy levels 1 and 2 within providers.

62. Design D does not reduce the cost of administering the survey or the costs to providers of complying with the requirement to take part. Its main advantage is that all students would be able to take part, providing insights into different stages of the academic experience and creating a shorter feedback loop between providers and students.

63. While design D would produce useful information at a national level, and add to our understanding of students' journeys, it does not reduce burden and adds some complexity. While we may wish to explore this option in the future, it is not recommended that we take design D forward at this stage.

Design E: Volunteer survey

Design

64. To run a volunteer survey, we would provide an online portal. Eligible students would authenticate themselves using personal details in order to respond. In its simplest form, there would be no target lists or sample selection: the sample would be determined by students' willingness to participate.[11] The volunteer survey would need to be supported by an effective promotion campaign, either coordinated centrally or by providers.

65. If running a volunteer survey, we would have no direct control over the response rates. For example, we could not reach out to students who have not yet responded, because we would not have any contact details. We do not know how many students would respond to a volunteer survey – this would depend on the effectiveness of the promotion campaign, and also whether the survey could inherit some of the momentum of the NSS. A large risk of this approach is that we might find ourselves without sufficient responses and unable to take remedial action.

66. In order to explore this design, we have assumed that 20 per cent of eligible students would respond (90,000 students). This assumption is based loosely on the number of responses made at present through the NSS portal, together with background knowledge about survey response rates. It may be either an under or an over-estimate.

Costs and burden

67. We estimate the financial costs of the survey to be relatively low – around 10 per cent of current costs. It is the cheapest of the survey designs. This is because the survey contractor would not need to contact individual students. This assessment includes a conservative estimate of the cost of a promotional campaign.

68. This design also has the potential to lower compliance costs for providers. In the simplest form of the design, they would not need to check target lists or supply contact details, as no direct contact would be made with students. It is possible that they would need to supply a student identifier for each eligible student; alternatively, the details used for authentication could be sourced directly from the administrative data collected by HESA and the ESFA. Providers could be required to promote the survey; alternatively, all promotion could be central. Understanding the estimates would be harder, because of the variable response rates and the possibility of bias (discussed below).

69. Fully volunteer surveys – that is, when any engagement is initiated by the respondent – are usually considered to be without respondent burden.

Accuracy

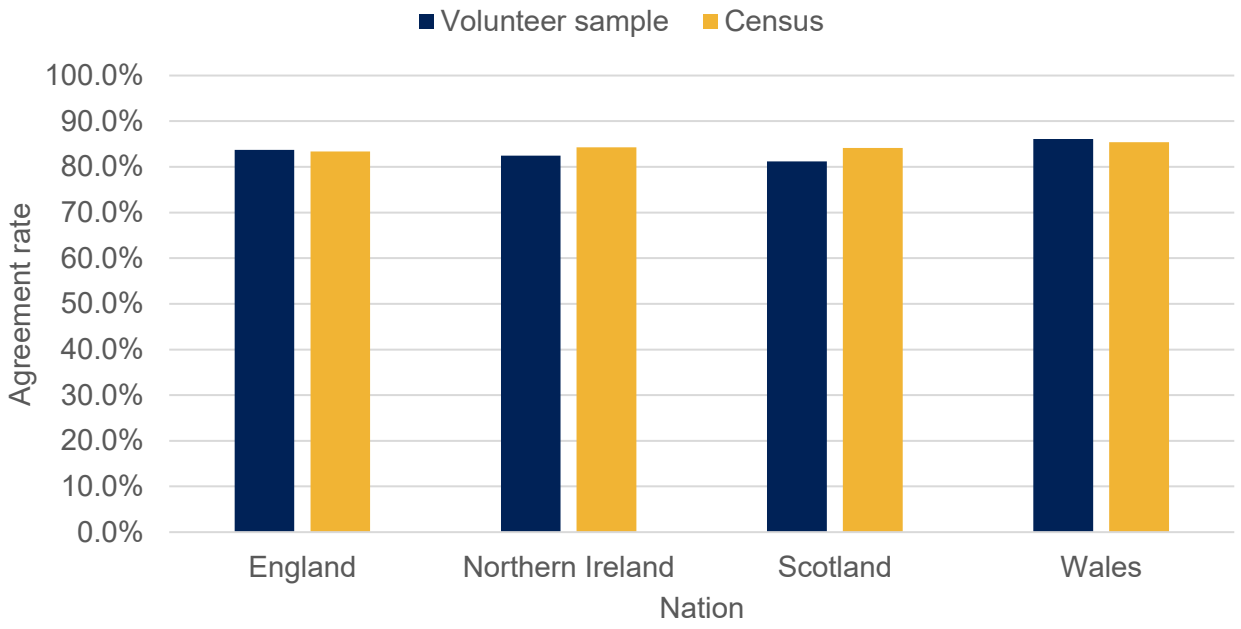
70. As a way of understanding the implications of design E, we evaluated what would have happened if, in 2019, we had received only online responses; and only 20 per cent of the

[11] A more complicated version of this design would allow providers to check and correct their target list of eligible students (the process currently known as additions and removals). This would reintroduce burden, but would ensure that all eligible students were able to respond.

eligible population had responded. To cancel out the effect of the different dates that providers chose to start NSS promotion, we normalised the time and date of response by reference to the first response the provider received. For example, if a provider's first response was made on 1 February and their next one on 2 February, we treated the second response as arriving on the second day of the survey. This approach does not directly simulate a volunteer survey. The earliest respondents are not necessarily those who would volunteer to complete the NSS without being targeted. But it does help us understand what happens when the target list is determined by some element of human decision making, rather than random selection.

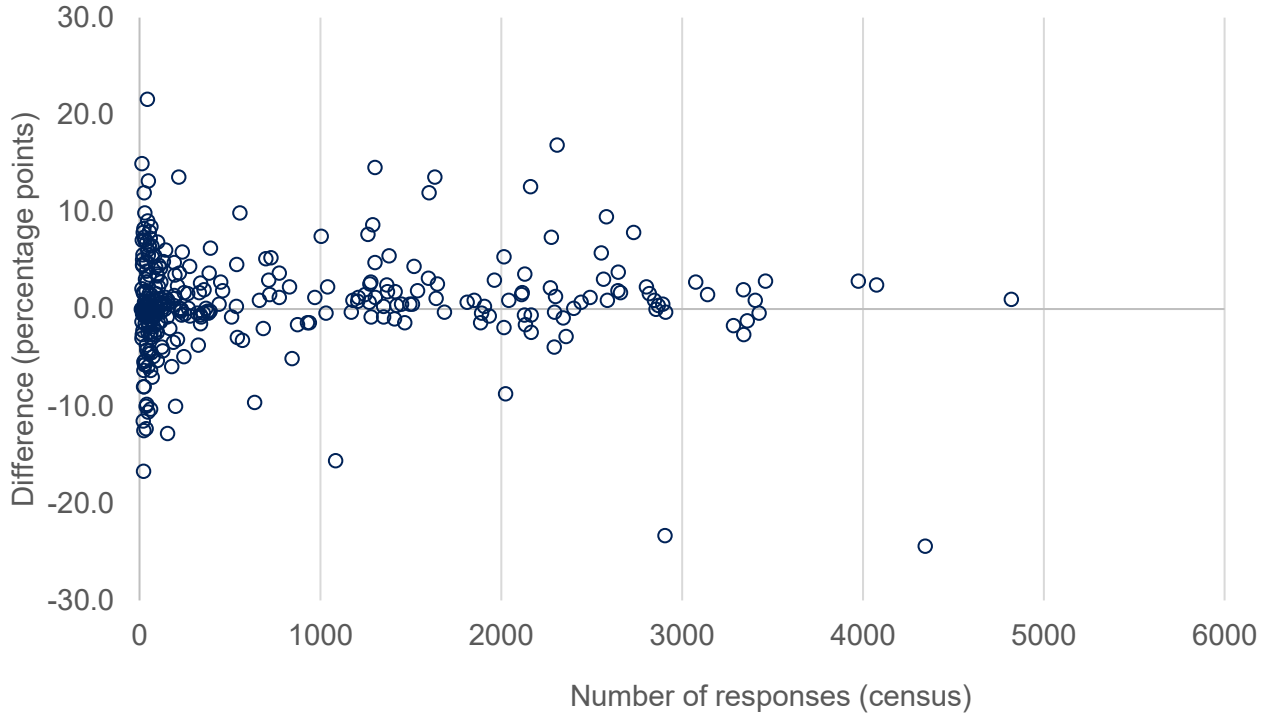
71. Figure 14 shows the Question 27 agreement rate derived from this simulation, compared with the estimates from the full census survey. Since we only have one version of the sample (rather than 100), space allows us to show the estimates for all four nations. We can see from the chart that the estimates from the simulation differ slightly from the estimates from the full census survey. This is not a surprising finding: we know from previous work that earlier respondents tend to respond differently from later respondents (and that this effect varies by nation).

Figure 14: Question 27 agreement rates by nation from design E simulation



72. Figure 15 shows the provider-level estimates from the design E simulation. Each dot is a provider. The Y axis shows the difference between the provider estimates and the estimate from the full census survey. This gives us a sense of the extent to which the results generated from a volunteer survey might be skewed. The estimates for some providers are close to the full census results. But for others the difference is greater than 10 percentage points. Estimates tend to be more positive, but some are more negative. The difference will be due to a mixture of sampling error and bias – that is, systematic distortion due to the survey design. If early respondents respond differently to later respondents in ways that matter, then a sample that includes only early respondents will yield results that do not accurately reflect the views of the whole population. We would expect the same to be true of a survey that included only volunteers.

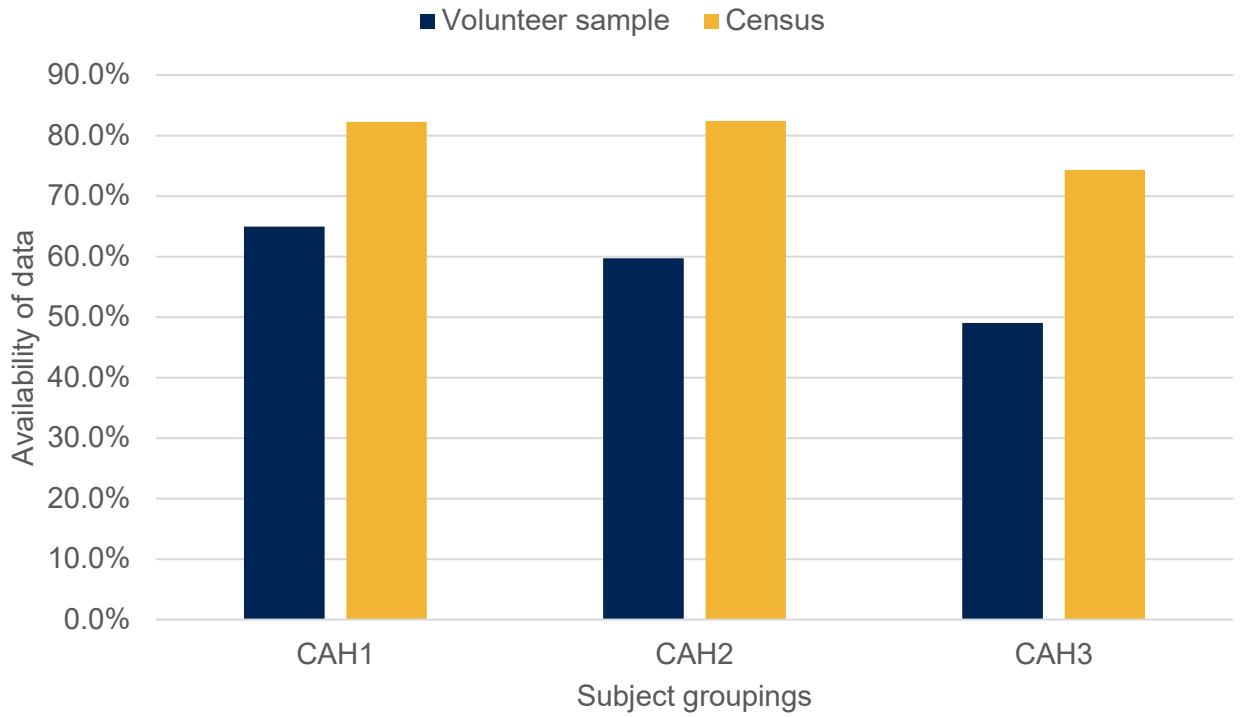
Figure 15: Difference in Question 27 agreement rates by provider from design E simulation compared to census



Availability of data

73. Figure 16 shows the availability of course level data from the design E simulation. When we use the high level CAH1 grouping, 65 per cent of subject groupings within providers meet the data availability threshold. When we use the more specific CAH3 grouping, this figure falls to 49 per cent.

Figure 16: Availability of data about courses (design E), compared with census survey



Design E conclusions

74. Design E is likely to be the cheapest option to implement. It also requires little or no engagement from providers. However, there are two related risks of a volunteer survey such as design E. Firstly, we would have little leverage over response rates. If response rates were low, or patchy, during a survey window we would not be able to contact students to encourage more responses. Secondly, the design does not depend on random selection, but instead on the willingness of students to proactively complete the survey. This introduces the possibility of bias: we form an incorrect view about the student experience because the students who responded to the survey are different from those who did not. Figures 14 and 15 give an idea of the possible scale of the bias.

75. We recommend that design E is rejected as a standalone option. However, it might be successfully combined with another survey design. For example, we could combine a volunteer survey with a large sample (such as design B), thereby giving all students the opportunity to respond, even if they were not selected in the sample. Running a volunteer survey alongside a random sample would allow us to understand, and take steps to mitigate, the bias.

Design F: Opinion poll

Design

76. Design F is an opinion poll. This would involve asking a polling company to survey students that they have access to – for example, through a standing panel that they maintain, or through contact details that they have purchased. This sort of poll is a form of convenience sample, in the sense that the students are selected because they are easy to access, rather than through any random or statistical mechanism. We have evaluated an opinion poll with a target of 10,000 students. This is relatively large by polling standards, but nonetheless gives us the smallest sample size of all our options.

Cost and burden

77. Based on the advice of three polling companies, we estimate that the cost of this design would be around 30 per cent that of the present survey.

78. The compliance costs for providers would be zero: providers would not need to do anything to allow the survey to take place.

79. Insofar as the respondents opt into the survey (for example, by agreeing to be part of a standing panel maintained by the polling company), the respondent burden would also be zero.

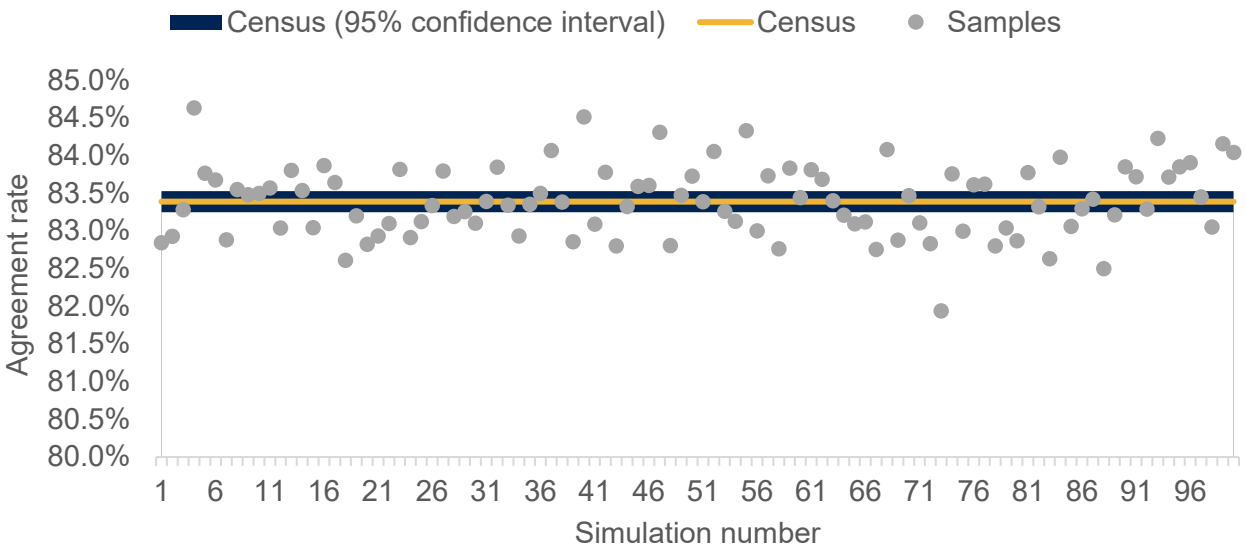
Accuracy

80. Like the volunteer survey, an opinion poll may be subject to bias. Students who volunteer to engage with polling companies may be relevantly different from other students. If so, the estimates from an opinion poll will tend to differ systematically from the truth about the target population. We have not attempted to quantify this bias, but note it as a real risk.

81. Like the other designs, the opinion poll will be subject to sampling error – that is, the estimates will vary from the true value simply because of the subset of students selected. To understand this, we have treated the opinion poll as a pure random sample of 10,000 students, and assessed the estimates we would expect from such a sample. This approach is stringent in some respects, and lenient in others. We would expect a polling company to put in place some form of quotas, which try to ensure that a balanced group of students are included in the sample. On the other hand, our random sample selects from the whole target population, whereas we would expect a polling company not to have access to some groups of students.

82. Figure 17 shows the Question 27 agreement rates for England generated using our very approximate simulation of design E. The likely range of estimates is from 82.8 per cent to 84.2 per cent, compared with a full census value of 83.4 per cent. Seventy-eight of the 100 estimates fall outside the confidence interval of the full census value.

Figure 17: Question 27 agreement rate estimates for England (Design F)



83. We have not assessed the reliability of provider-level estimates from design F because we do not expect these estimates to be available: see paragraph 84 below.

Availability of data

84. We do not expect an opinion poll to provide information about the student experience at individual providers. Even if the polling company had access to students at every provider, a sample size of 10,000 is simply too small to support estimates at this level. Similarly, an opinion poll would not provide data about individual courses.

Design F conclusions

85. An opinion poll such as design F might allow us to detect large changes at the national level. It would not allow us to confidently detect smaller changes – for example, those of one or two percentage points – both because of sampling error and the possibility of bias. An opinion poll could not be used to understand the student experience at individual providers, or to provide information about courses. Since the design does not provide data that meets the needs of users, we do not recommend taking it forward.

Annex A: Needs of data users

Funders and regulators

1. The fundamental unit at which most regulation and distribution of funding occurs across the UK higher education sector is the provider. However, there are also regulatory and funding processes that make use of NSS data for subject areas within providers and it is used to examine national trends too. We have identified the following uses of data by the funders and regulators of higher education as part of this review.

Table 1: Uses of NSS data by UK funders and regulators

Data user	Level of data	Comments
Office for Students	Nation	To inform policy and to measure the OfS's performance.
	Provider	Currently used in the Teaching Excellence and Student Outcomes Framework (TEF). The current consultation on regulating quality and standards in higher education (see Annex C)[12] proposes the use of data from national student surveys or from student polling for monitoring compliance of individual providers with their conditions of registration, and references the NSS as an example of data that could be used for this purpose.
	Demographic and subjects within nation	Currently used to inform policy decisions and to provide context to other measures.
	Demographics and subjects within providers	Currently used in TEF to inform assessment; may be used in the future to inform quality and standards.
HEFCW	Provider	Regulating quality and risk review. Governing bodies at providers are required to review and act on NSS results.
	Subjects within providers	As above
Department for Education (NI)	Provider	Assessing quality and standards
Scottish Funding Council	Nation	Accounting for use of public funding to Scottish government
	Provider	As key part of Quality Enhancement Framework
	Subjects within providers	As above

[12] For information about the consultation, see: www.officeforstudents.org.uk/publications/consultation-on-regulating-quality-and-standards-in-higher-education/.

2. In summary, the use of NSS data by funders and regulators to assess quality and to drive enhancements requires data to be available both at provider and subject-within-provider levels. Some also use national level figures for informing policy and providing public accountability. These then seem to be the most important levels of data any survey method should accurately produce, though it should be noted that other population and within-provider breakdowns (for example, such as disabled students within a provider) remain important. For simplicity, we have not discussed these other breakdowns in our evaluation, but further information is available on request.

Students, including prospective students

3. User testing research carried out on behalf of the OfS as part of the Unistats and Discover Uni website development process has consistently found that prospective students are interested in highly granular data. They want to understand the experience of students on the specific courses that they are thinking of applying to. They are less interested in the student experience at provider level, or for nations as a whole. This is borne out by the student polling carried out for this review which showed that 57 per cent of respondents thought course-level data about students' academic experience was "extremely useful". User testing for the Discover Uni website shows that, though prospective students prefer data about specific courses, they are willing to accept subject-within-provider data if it is the best available option. Given this, we have focused on subject-within-provider level data to assess the usefulness of each method for student information purposes. Although this is not as granular as the current information available, it avoids adding an additional criterion at a level of detail not required by funders and regulators while still providing a useful indication.

Providers

4. Like students, providers tend to value course-level data. As part of our review, we asked interested parties whether NSS data would be less useful for improving the student experience if it were not available at course level. Seventy-six per cent of those responding on behalf of a provider agreed this was the case. We also know that providers make use of course-level data from the queries we receive. These often relate to the results for a particular course, or subject grouping; or requests for even more granular data. Providers also use provider and provider-subject level data in their marketing, and as a way of comparing themselves with other providers.

General public and third party users

5. NSS results are of interest to the general public, as demonstrated by the yearly interest from journalists in the release of the data and the use of the data by compilers of league tables and private student information services. Although the use of NSS data for some of these purposes may play a role in some negative processes, such as gaming or grade inflation, this needs to be weighed against the public interest in this data. These uses of the data however reflect the same interests as those of funders and regulators: nation, provider and subject-within-provider being the most interesting data items for these users.

Summary

6. Data users have a variety of preferences and requirements when considering NSS data. The minimum based on the above seems to be that useable data must exist at nation, provider, and subject-within-provider levels to meet the most essential needs of the data users. Course-level information is mainly used by providers and by prospective students; in both cases the subject-within-provider level provides a close proxy, so we have not made this an essential criteria for our consideration. For these reasons we have focused our analysis on nation, provider and subject-within-provider for the purpose of this report.