R11 Denbigh 2 Friday 7 December 9.00 - 9.30

What proportion of National Student Survey comparisons are statistically significant? (0366)

Alex Buckley[1] [1]*University of Strathclyde, Glasgow, United Kingdom*

1. **Introduction**

The National Student Survey is a powerful force in UK higher education. Comparisons of NSS results are intended to play a key role in prospective students' choice of institution, though there is evidence that this effect may be limited (HEFCE 2014, Gibbons et al 2015). There is growing evidence that comparisons of NSS results play a key role in quality assurance and enhancement within institutions (Buckley 2012, Gibbs 2012, Pickford 2013). Despite these influences the NSS has been under-researched (Ashby et al 2011, Lenton 2015), and as a consequence data is often used simplistically. Within institutions, firm conclusions are drawn from small differences in raw scores, without consideration of the statistical properties of the data. This is not a problem limited to the NSS, as evidence suggests that student feedback data is often over-interpreted (Abrami 2001, Boysen et al 2014).

Previous research has highlighted the limitations of the NSS when making comparisons (Surridge 2009, Cheng and Marsh 2010), and from the outset the aim of generating results at a sufficient fineness of grain that are still reliable was acknowledged as "ambitious" (HEFCE 2004a). The importance of statistical significance for the presentation of NSS results has been noted numerous times over many years. While it was originally proposed that 95% confidence intervals would be included on the public presentation of the data (HEFCE 2004b), the primary mechanism for the public presentation of the results (the Teaching Quality Information website and subsequently Unistats) has not included information about statistical significance. The consultation on the revisions to the NSS in 2015 again highlighted the need for better explanation of the effect of sample sizes on comparisons of the data (HEFCE 2015), and researchers made the same point (Cheng and Marsh 2010).

HEFCE responded to these concerns in 2012 by creating institutional benchmarks employing indicators of significance, and this idea has been transferred to the Teaching Excellence Framework, where an institution's 'flags' are determined by whether the differences between their NSS results and their benchmarked score are statistically significant. Nevertheless, when NSS results are used for comparisons on Unistats or when presented in the media it is raw scores that are used with little consideration of the statistical limitations. Institutional use for quality enhancement and assurance also seems to depend heavily upon raw scores. Given this reliance on raw scores in making comparisons using the NSS, it is important to explore the impact of the statistical limitations of the data on the reliability of those comparisons.

Method

This study uses methods developed in previous work designed to support institutions in making sense of their own NSS data for quality enhancement and assurance (Buckley 2018). Those methods draw on the 95% confidence intervals provided with the publicly available

aggregate NSS data (representing the proportion of students agreeing with each NSS statement) in order to assess the statistical significance of common kinds of comparisons. Differences are judged statistically significant (at the 99% level) if there is no overlap in the (95%) confidence intervals provided by HEFCE (Cumming and Finch 2005).

The paper aims to answer the following three questions:

1. What proportion of institutional NSS scores differ statistically significantly from the sector average?
2. What proportion of institutional NSS scores differ statistically significantly from year to year?
3. What proportion of comparisons between institutional NSS scores are statistically significant?

All of these questions are explored at the subject level.

Results

In 2017 there were 135,999 institutional NSS scores at the JACS3 level of subject grouping (with 108 subjects) for the 27 items in the questionnaire. Overall, 23.8% of these were found to differ statistically significantly from the sector average (for the relevant subject and question). A university department can therefore expect their scores for around a quarter of the NSS questions to differ significantly from the sector average. This figure differs by subject, from 66% for Veterinary Sciences, to 3.8% for French.

Investigation of year-to-year changes from 2013 to 2016 (prior to a change in the questionnaire wording) found that overall, 2.1% of the year-to-year changes were statistically significant. On average therefore, a university department could expect that fewer than one of their 22 NSS scores will have changed in any one year. Again, this differs by subject. It also differs by NSS item, with item 7 - on the promptness of feedback - the most likely to change meaningfully (5%) and item 20 - on the improvement of communication skills - the least likely (0.7%).

For the exploration of the significance of comparisons between institutions, definitive findings will be presented in the paper but only preliminary results are available at present. Based on a representative sample of 10 institutions, and using the JACS1 subject grouping consisting of 18 subjects rather than the more fine-grained JACS3 grouping, it was found that 9.6% of the 20,808 comparisons between institutions were statistically significant. This differs markedly by subject, from 20.7% for Creative Arts to 2.1% for Engineering and Technology. It also differs by question, from 31.7% for question 15 - on the organization of the course - to 2.4% for question 21 - on feeling part of a community of staff and students. For overall satisfaction (question 27) the figure is 11.8%, so slightly more than one in ten of the institutional comparisons are statistically significant.

Conclusion

Given its intended role in student choice and its actual role in institutional quality mechanisms, it is important to ascertain the ability of NSS results to meaningfully compare

institutions, to indicate changes over time, and to distinguish institutions from the sector average. This paper will present findings indicating that the ability of the NSS to perform those roles is limited: for the vast majority of subjects and questions, institutions do not differ statistically significantly from the sector average, or from more than a small minority of other institutions, and year-on-year changes are very unlikely to be statistically significant.

References

Abrami, P. (2001) 'Improving judgements about teaching effectiveness using teacher rating forms', *New Directions for Institutional Research* 109: 59-87

Ashby, A., Richardson, J. and Woodley, A. (2011) 'National student feedback surveys in distance education: An investigation at the UK Open University', *Open Learning* 26(1): 5-25

Boysen, G. (2017) 'Statistical knowledge and the over-interpretation of student evaluations of teaching', *Assessment and Evaluation in Higher Education* 42(7): 1095-1102

Buckley, A. (2012) *Making it count: Reflecting on the National Student Survey in the process of enhancement* (York, HEA)

Buckley, A. (2018) 'How to use the National Student Survey (responsibly)', presentation at the *Enhancement Themes Conference*, Glasgow Caledonian University, 7 June 2018. Available at: http://www.enhancementthemes.ac.uk/docs/ethemes/evidence-for-enhancement/how-to-use-the-national-student-survey---responsibly-(paper).pdf

Cheng, J. and Marsh, H. (2010) 'National Student Survey: Are differences between universities and courses reliable and meaningful?' *Oxford Review of Education* 36(6): 693-712

Cumming, G. and Finch, S. (2005) 'Inference by eye: Confidence intervals and how to read pictures of data', *American Psychologist* 60(2): 170-180

Gibbons, S., Neumayer, E. and Perkins, R. (2015) 'Student satisfaction, league tables and university applications: Evidence from Britain', *Economics of Education Review* 48: 148-164

Gibbs, G. (2012) *Implications of 'Dimensions of quality' in a market environment* (York, HEA)

HEFCE (2004a) *National Student Survey 2005: Outcomes of consultation and guidance on next steps* (Bristol, Higher Education Funding Council for England)

HEFCE (2004b) *National Student Survey 2005: Consultation* (Bristol, Higher Education Funding Council for England)

HEFCE (2014) *UK review of the provision of information about higher education: Advisory study and literature review* (Bristol, Higher Education Funding Council for England)

HEFCE (2015) *Review of information about learning and teaching, and the student experience: Consultation on changes to the National Student Survey, Unistats and information provided by institutions* (Bristol, Higher Education Funding Council for England)

Lenton, P. (2015) 'Determining student satisfaction: An economic analysis of the National Student Survey', *Economics of Education Review* 47: 118-127

Pickford, R. (2013) 'Leadership of the National Student Survey for enhancement', *Quality Assurance in Education* 21(4): 344-358

Surridge, P. (2009) *The National Student Survey three years on: What have we learned?* (York, Higher Education Academy)