

Foreword

The Higher Education Academy has undertaken and commissioned a range of research on data arising from the UK National Student Survey. This synthesis report provides a brief overview of some of the findings arising from research at sectoral level, identifies the questions raised by that work and suggests areas for future research.

The NSS has many strengths, particularly in helping us to understand more about student experience of undergraduate education, but also has some limitations in this regard. This report not only outlines the areas in which the NSS can be a useful data source, but also clarifies what it should not be used for.

The enhancement potential of NSS is quite often forgotten in the league table haze and many institutions find themselves preoccupied with the machinery of the annual survey, to the detriment of careful analysis of what the survey data can illuminate. The Academy is uniquely positioned in the HE sector to support use of NSS for enhancement purposes and, to that end, we have worked with a number of colleagues and stakeholders in the sector in a range of ways.

For example, we facilitate the NSS working group: NSS administrators from a number of institutions who share and develop good practice in using NSS data for enhancing students' learning and teaching experience. A number of institutional case studies can be accessed on our website and will soon be accessible through our EvidenceNet. The Academy has hosted the Annual NSS conference, which brings together colleagues from across the sector in order to share research data and effective practice of using survey evidence to inform policy and practice. The event is now broader in its scope to allow exploration of data arising from a range of surveys of student experience – for example, the Postgraduate Research Experience Survey (PRES) and Postgraduate Taught Experience Survey (PTES) – in addition to the NSS.

We are also fortunate in having access to the full NSS dataset, on the basis of which we commissioned Herb Marsh, Antony Fielding and James Williams to conduct analyses that considered specific parameters of NSS data (i.e. degree attainment, assessment and feedback issues). We have also published guides to multilevel analysis (to complement tables included in Surridge reports commissioned by HEFCE) and to the bank of NSS optional items.

Moreover, there is a lot of interesting NSS work taking place in the Academy at disciplinary level. The Economics Subject Centre triangulates their own student experience survey with NSS results at subject level and provides departments with confidential reports. The Subject Centre for Sociology, Anthropology and Politics (C-SAP) ran an event to look closely at NSS results for Politics and triggered a discussion on the possible reasons why Politics scores differently (from other disciplines) on some NSS scales. The Art Design and Media Subject Centre commissioned Mantz Yorke to investigate undergraduates' experiences in Art and Design by triangulating NSS results with First-Year Experience Survey results.

NSS helped to identify whole enhancement themes like assessment and feedback, and personal tutoring, both of which are strands of the Academy's work.

The Academy will continue to play a key role in surfacing and synthesising findings arising from NSS research and sharing effective practice in the utilisation of NSS results, to enhance students' learning experiences. Core tasks for the future include provision of support to FE colleges that took part in NSS for the first time in 2008, supporting and developing the NSS working group and hosting the Survey conference in Spring 2009.

This synthesis report has been commissioned from Paula SurrIDGE, whose HEFCE-funded research into NSS data has made an important contribution to the understanding of this instrument. We are delighted that Paula has completed this work for us.

I would like at this stage to acknowledge the contribution of Gosia Kulej whose work at the Academy on a wide range of surveys has also made a significant impact in this field.

I hope that this synthesis report will be of use to a variety of colleagues in the sector and will help to debunk some myths, provide some food for thought and illuminate some possibilities around interpretation and use of NSS data.

Rachel Segal, Assistant Director, Research and Evaluation
The Higher Education Academy

The National Student Survey three years on: What have we learned?

Paula Surridge

Department of Sociology
University of Bristol
email: p.surridge@bris.ac.uk

Table of contents

Executive summary.....	5
1. Context	7
1.1 The birth of the NSS	7
1.2 The NSS evolved.....	8
2. Data and methodology.....	8
2.1 The structure of the NSS	8
2.2 The questionnaire	9
2.3 Other data.....	10
2.4 Response rates and publication thresholds	10
2.5 Analysis techniques	13
3. Key findings from NSS research.....	13
3.1 Surridge: Reports on 2005, 2006 and 2007 data	14
3.2 Fielding <i>et al.</i> : Assessment and Feedback, ethnicity and gender.	22
3.3 Marsh and Cheng: Dimensionality and multilevel structure	24
4. Synthesis: Three years of NSS research	26
4.1 Synthesis	26
4.2 Future research	27
4.3 The NSS: Strengths and limitations	28
5. The NSS: What does it mean for us?.....	29
5.1 Potential students and the NSS	29
5.2 Institutions and the NSS	29
5.3 The HE sector and the NSS.....	30
6. Published research and further information	31
Published reports.....	31
Other sources of information.....	31
Appendix: Multilevel models	32

Executive summary

- This report brings together the key findings of three years of research using the NSS. It aims to provide a non-technical summary of these key findings, a guide to interpreting this work and a discussion of the strengths and limitations of the NSS as a vehicle for understanding the student experience.
- The report summarises three key pieces of research: the analysis of the 2005, 2006, and 2007 NSS by Surridge, published by HEFCE, which focuses on the effects of student, course and institutional characteristics on the student experience; an analysis of gender and ethnicity effects on Assessment and Feedback by Fielding *et al.* and published by the Equality Challenge Unit and the Higher Education Academy (the Academy); and research by Marsh and Cheng, also published by the Academy, which focuses on the dimensionality and multilevel structure of the NSS data. The report also includes findings from previously unpublished research relating to response rates and incentives offered by institutions.
- Institutional response rates ranged from 20.3% to 84.2% in 2005, 16.0% to 80.5% in 2006 and 41.8% to 80.1% in 2007. This variation in response rates among institutions has raised questions about the validity of comparisons between institutions with very different response rates. However, under the assumption that non-respondents to the survey have experiences that are similar to 'late' responders to the survey, there is no evidence to suggest that scores would be materially different given higher response rates.
- In 2006 and to a greater extent in 2007, some institutions offered incentives to students to respond to the NSS. In total 12 institutions offered such incentives in 2006, increasing to 52 institutions in 2007. Using multilevel models of the likelihood of a student responding to the NSS in 2006 and 2007, no statistically significant impact of incentives on the likelihood to respond was found.
- The reports by Surridge highlight both the complexity and stability of student, course and institutional characteristics on the student experience. They conclude that no single characteristic has the same effect on all aspects of the student experience; however, where effects are found on any given aspect of student experience, these are, for the most part, stable over the three years of the survey.
- The report by Fielding *et al.* highlights the complexity of differences according to ethnic group on one aspect of the student experience: Assessment and Feedback. A general conclusion is that Black students tend to be more positive about Assessment and Feedback than White students, while students from Asian backgrounds were less positive about Assessment and Feedback.
- Marsh and Cheng suggest that two approaches could reasonably be used to generate a summary of overall satisfaction, these are either a summation of the individual items (which should be weighted to take into account the different contribution each makes to overall satisfaction) or the single Overall Satisfaction measure (question 22); the latter approach is less appropriate as

a feedback mechanism as it is important for institutions to also know which elements that comprise satisfaction are relative strengths and weaknesses.

- Both Surridge and Marsh and Cheng find that the proportion of variation in student experience that is attributable to institutions is small, but statistically reliable. The question of the substantive significance of these differences remains open to interpretation. One interpretation is that these differences are too small to be of use to potential students in forming decisions about where to apply for courses, the other is that the small differences should give confidence to students that they will receive a generally good experience of higher education wherever they choose and that choice may be guided by other factors, such as disciplinary specialisation. The NSS data cannot discriminate between different interpretations.
- Both Surridge and Marsh and Cheng conclude that there are large differences according to subject of study, and suggest the need for caution when making comparisons either between different subject groups within an institution or within a subject group across institutions.
- Three elements of the research summarised in this report are key for institutions:
 - First, the need to contextualise NSS scores on the basis of the types of students at an institution. The research has shown that different groups of students have different experiences of HE, and institutions should aim to be sensitive to the experiences of the types of students that make up its student body.
 - Secondly, the research provides a contextualised measure of institutional performance via the institutional residuals. These residuals are an important supplement to on-the-ground knowledge about different aspects of teaching quality at an institution. They provide a measure of performance not directly available to institutions as they show performance relative to the rest of the HE sector.
 - Finally, the research suggests that the differences between institutions are stable and reliable. However, it also shows that these differences are small, either as measured by the number of institutions that have statistically significant residual terms or as measured by the proportion of variation explained by differences between institutions. The vast majority of students rate their higher education experiences positively, and the vast majority of institutions are not statistically different from each other in this regard.

1. Context

The National Student Survey (NSS) not only provides invaluable information to prospective students and institutions about student experiences of higher education in the United Kingdom, it is also an incredibly rich resource for research into the student experience. This report brings together the key findings of three years of research using the NSS. It aims to provide a non-technical summary of the key findings, a guide to interpreting this work and a discussion of the strengths and limitations of the NSS as a vehicle for understanding the student experience. While the report will be of use to advisers of potential students and the wider policy sector, it is primarily aimed at those working within HE institutions, for whom the NSS provides a valuable source of data about the student experience within their own institutions. However, a key strength of the NSS is the ability to make comparisons across the sector as a whole; the research discussed in this report should be seen as providing an additional resource to institutions with regard to a contextualisation of their own results in the light of results elsewhere in the sector.

1.1 The birth of the NSS

The first full National Student Survey (NSS) was conducted in 2005 at higher education institutions in England, Wales and Northern Ireland. It has been conducted on an annual basis in subsequent years, with some institutions in Scotland opting into the survey in 2006 and 2007. The NSS was developed as part of the revised Quality Assurance Framework for higher education, which came about at the end of subject review as conducted by the Quality Assurance Agency for Higher Education (QAA). The original briefing document¹ states that the aims of the NSS are:

1. to inform the choices of future students, alongside other sources of information about teaching quality
2. to contribute to public accountability by supporting external audits of institutions by the QAA.

Thus, from its inception the NSS has had multiple purposes and been required to inform multiple audiences. To satisfy the first of these purposes data from the NSS are made publicly available as part of a package of information on teaching quality, currently accessed via the Unistats website (www.unistats.com). In addition, institutions have more detailed access to their own NSS results than those published via the website, as concerns about reliability and privacy mean that publication thresholds must be met before data are made publicly available (currently these thresholds are set at a 50% response rate and 23 respondents for any given unit under examination²). These publication thresholds apply to all NSS data made available via the Unistats site,³ but do not apply to the data given back to institutions on an individual basis or to the research discussed in this report.

¹ *National Student Survey 2005: Outcomes of consultation and guidance on next steps.* Available from: www.hefce.ac.uk/pubs/hefce/2004/04_33.

² See Section 2.4 for more discussion of the impact of the thresholds.

³ Data provided directly to institutions are subject to a threshold of ten respondents per cell.

1.2 The NSS evolved

The data from the NSS are now used in a vast array of arenas, both publicly and within individual institutions: to measure the quality of teaching in institutions for the purposes of improvement within institutions, for audit purposes and for the purposes of 'ranking' institutions according to their scores on the survey. As with any large body of public data, not all the uses to which the data are put are reliable, some are not desirable and many are not within the original expectations of the survey. In the context of the NSS, the data are increasingly used by the media to feed into league tables and by institutions as a marketing tool, as well as by those for whom it was originally intended: prospective students and internal quality assessment processes. In many cases the data may be used in ways that are not entirely appropriate or used to draw conclusions that are not founded. This use of the data rightly causes concern within the HE sector; however, the key aim of this report is not to debunk the uses of the data made in the media, but rather to aid an appropriate understanding of the data and awareness of current research findings among the HE sector itself.

2. Data and methodology

2.1 The structure of the NSS

In order to provide comprehensive coverage of subject areas and reliable data for comparisons between units, the NSS is a census of students in their final year of a course leading to undergraduate credits or qualifications across the UK. In addition, the same basic questionnaire is administered to all students to ensure comparability between institutions and subjects.⁴

The survey fieldwork has been conducted by IPSOS-MORI and has followed the same basic structure each year. The survey fieldwork begins during January, with the fieldwork start date staggered by institution. Regardless of the start date, each institution follows the same timetabled structure of contacts with students as shown in Table 1.

⁴ While the same basic questionnaire has been used in all years, there have been additional questions used with some specific groups of students (e.g. NHS-funded students) and also additional optional questions in which institutions could choose to participate. The responses to these optional questions are not made available via Unistats, rather they are intended to provide additional information about teaching quality to the institutions that opt in to them.

Table 1: The student contact process

Week 1	Students emailed with a link to the NSS website inviting them to respond online. Reminder emails sent one week later to non-respondents
Weeks 2-3	Paper questionnaires distributed to non-respondents, reminder postcard sent one week later
Weeks 4-8	Telephone contact with non-respondents

At any stage of this process students may respond by visiting the NSS website, so while there is a strong patterning of response type by time of response, it is not the case, for example, that all responses after Week 4 of the survey are received by telephone.

2.2 The questionnaire

The basic NSS questionnaire is made up of 21 items, designed to capture six essential dimensions of teaching quality, and an additional item to measure 'overall satisfaction'. Students are also invited to add further qualitative comments about their experiences in two open-ended questions, which ask them to highlight particularly positive and negative aspects of their experiences. These qualitative comments are fed back to institutions, but are not made available via the public website. Our concern in this report is on analysis of the 22 quantitative questions that form the core NSS questionnaire. For the purposes of analysis these 22 items are grouped together to form 'analytical scales', that is groups of items measuring the same underlying aspect of teaching quality are used together to provide a more reliable measure of the quality of that specific area.

The scales are:

- Teaching and Learning (questions 1-4)
- Assessment and Feedback (questions 5-9)
- Academic Support (questions 10-12)
- Organisation and Management (questions 13-15)
- Learning Resources (questions 16-18)
- Personal Development (questions 19-21).

Each of the individual question items has five possible response categories,⁵ with each response having a numerical score allocated to it:

5. Definitely agree
4. Mostly agree
3. Neither agree nor disagree
2. Mostly disagree
1. Definitely disagree.

⁵ It is also possible for a student to respond 'Not applicable'.

The scales are created by adding together the scores for the responses given by the student to each of the scales' constitutive items; this total is then divided by the number of items that make up the scale to restore the scale to the original measurement values. Thus, each of the scales takes a value between 1 and 5.

The final item (question 22), which measures 'overall satisfaction', is analysed as a single item.

2.3 Other data

The research discussed in this report uses data collected via the NSS questionnaire linked with a student identifier to the student HESA records. This enables the researcher to look in detail at different types of students and their NSS responses, as well as having characteristics of courses, such as subject of study. This data linkage is provided to the researchers by HEFCE and is not available to individual institutions or via dissemination websites (in order to protect the anonymity of students). This means it is not usually possible for individual institutions to conduct the type of analyses discussed in this report.

2.4 Response rates and publication thresholds

In order to ensure the anonymity of students, while also protecting data quality, data for individual institutions and subject areas within them are only made available publicly where a minimum threshold is met. Currently this publication threshold is set at a 50% or greater response rate and 23 or more students. This threshold ensures a minimum level of data quality across institutions; however, variation in response rates exists between institutions.

Response rates

The overall response rate was around 60% in 2005 and 2007, although it showed a slight dip in 2006 to 57%, which was due in large part to changes in the number of times a student was contacted by telephone to elicit a response. However, these generally good response rates for the sector as a whole do mask a degree of variation in the response rates for specific institutions and also for specific groups of students.

Institutional response rates ranged from 20.3% to 84.2% in 2005, 16.0% to 80.5% in 2006 and 41.8% to 80.1% in 2007. This variation in response rates among institutions, in particular, has raised questions about the validity of comparisons between institutions with very different response rates. This issue is only of importance if the students who did not respond to the survey have substantively different experiences of higher education than those who did respond. There is no simple way to assess whether this is the case. One potential method would be to try to contact non-responders to elicit their responses; however, such an exercise would both be costly and also risk alienating the student body via further follow-ups to those who chose not to respond. Therefore, an alternative way of assessing the potential impact of this was sought that made use of the existing NSS data, and therefore placed no additional burden on students.

Late responders

In survey methodology it is often assumed that survey response follows a continuous scale related to time of response, with those who respond later in the fieldwork period more like non-responders than those who responded at the beginning of the period. Using this as the basic methodology, it is possible to use the timing of a student's response to the NSS to assess the extent to which those who responded later in the process had different experiences of HE than those who responded earlier in the process.

In previously unpublished work, Surridge used this assumption to test if late responders to the NSS were different to early responders⁶ regarding both the characteristics of late responders and their experiences of higher education. In this context, a 'late' responder was defined as a student who fell into the latest 10% of students to respond. This was defined on an institution by institution basis⁷ to avoid over-recruitment of late responders from specific institutions.

Using this measure in a multi-variate analysis, which included a range of student, course and institutional characteristics, showed that the effect of being a 'late' respondent on the analytical scales and on Overall Satisfaction was small and in some cases not statistically significant. Where the effect was statistically significant the difference between 'late' respondents and all other respondents was less than 0.05, in other words less than the margin of error produced by the rounding of figures to two decimal places.

The conclusion of this analysis is that, assuming non-respondents to the survey have experiences that are similar to 'late' responders to the survey, there is no evidence to suggest that scores would be materially different given higher response rates.

Thresholds

The converse of the analysis discussed above relates to the publication thresholds based on response rates. Here the key question, rather than whether the results would be different if response rates were higher, is whether results would be different if response rates were lower. In order to assess this, a similar methodology was applied, but in this case the response rate for each institution was artificially reduced to 40% by selecting only the first 40% of responses within each institution (and again within each method of response). This analysis shows that at the aggregate, sector-wide level, the impact of a reduction in response rate to 40% was negligible.

However, the key issue is how this would affect the publicly available data on the Unistats website. The Unistats site reports the '% agree' figures for each of the JACS2⁸ subject groupings within institutions. Based on the Overall Satisfaction measure (question 22), Table 2 shows the impact of reducing response rates to 40%

⁶ The analysis of 'late' responders and thresholds were conducted using the 2006 NSS data, by Paula Surridge at the University of Bristol on behalf of HEFCE.

⁷ It was also calculated within each method of response, that is the latest 10% of online respondents, the latest 10% of telephone respondents and the latest 10% of postal respondents within each institution. However, analysis confirmed that the substantive conclusions were the same if the 10% was applied to respondents as a whole or within method of response.

⁸ Joint Academic Coding System - this coding system is used by the Higher Education Funding Council to group together courses in cognate subject areas.

for each of the JACS2 subject groups within institutions. The measures used are directly comparable to those publicly available on the Unistats website. In each case, the actual '% agree' for all respondents was compared with the notional '% agree' for the 40% response rate.

Table 2: Summary of effect of reducing sample to 40%, within institution and JACS2 code

Total units	2,807	
Unreportable due to n < 23	1,376	
Total reportable units	1,431	100%
Change < 1%	447	31.2%
Change 1-3%	587	41.0%
Change 3-5%	239	16.7%
Change > 5%	158	11.0%

Source: Previously unpublished research

In total over 70% of the reportable units saw the '% agree' figure change by less than three percentage points as a result of a reduction in sample size to 40%. Change of more than five percentage points occurred in 11% of the reportable units. While in statistical terms these differences are small, many of the differences between institutions are themselves small. Therefore, a reduction of the response rate for institutions is unlikely to substantially affect the NSS scores for that institution, but as these measures are very sensitive to small changes it may affect the rank ordering of institutions overall.

Incentives

In 2006 and to a greater extent in 2007, some institutions offered incentives to students to respond to the NSS. The incentives varied widely in form, from free hot drinks, through printer credits, to prize draws to win iPods, book tokens and other items. In total, 12 institutions offered such incentives in 2006, increasing to 52 institutions in 2007.

While there is substantial variation in the nature of the incentives offered to students, the presence of any form of incentive indicates a particularly high level of institutional commitment to the NSS, and in particular to achieving a good response rate to the survey. However, is there any evidence that such incentives do indeed produce higher response rates?

A previously unpublished piece of work conducted by Surridge suggests that there is in fact very little impact on response rates from institutional incentives. Using multilevel models of the likelihood of a student responding to the NSS in 2006 and 2007, no statistically significant impact of incentives on the likelihood to respond was found. However, it was not possible to test whether particular types of incentives were more effective than others due to the relatively small numbers of institutions offering any kind of incentive. In conclusion, it is likely that the effect of incentives,

per se, is small and not statistically significant; however, it is possible that some particular types of incentives may be effective in improving response rates. In order to assess this, further research that involves working more directly with students (perhaps via focus groups) would be required.

2.5 Analysis techniques

The research reports discussed below use multi-variate statistical techniques in order to take account of a wide range of factors that may potentially influence students' responses to the NSS. Multi-variate techniques allow for a number of factors to be taken into account at once. For example, in a simple analysis we may find that female students and students on courses in Languages were more positive than some other groups. However, we cannot say whether the more positive scores for one might be connected to the other. In other words are those on courses in Languages more likely to be female and therefore more likely to be positive about their experiences? Multi-variate techniques allow us to assess if this is the case by including a range of relevant factors at the same time.

As well as the need to take into account many student characteristics at once, the data have an inherent structure, where students are clustered together within institutions. We might expect students at the same institution to have more common experiences than students at different institutions. In statistical terms, this means that the data for two students at the same institution may be correlated. This requires specialist models that take this correlation into account; these are known as 'multilevel' models. Multilevel models are used in different ways in the research reports discussed below. To aid in understanding what these models can achieve, a short introduction to the models is presented in the Appendix.⁹

3. Key findings from NSS research

The NSS provides an incredibly rich resource for research into the student experience in the United Kingdom. It is possible to use the data in many ways to approach a range of research questions. The research summarised in this section has begun to open up these research questions, but is by no means exhaustive in this regard. Additional research agenda that could be pursued with the data are discussed in Section 4.2 below.

This section summarises the key findings from published research reports using the NSS data. There are five main reports: firstly three reports by SurrIDGE (published by HEFCE) that cover the NSS data from 2005, 2006 and 2007; these reports are particularly concerned with variations according to student, course and institutional characteristics, as well as sector-wide institutional variation. Secondly, a report by Fielding *et al.*, published by the Equality Challenge Unit and the Higher Education Academy, which, although primarily concerned with differences in degree attainment,

⁹ Also a useful non-technical guide on multilevel analysis can be downloaded from the Academy website:
http://www.heacademy.ac.uk/resources/detail/ourwork/research/NSS_interpreting_data_using_multilevel_modelling.

contains a detailed examination of issues of assessment and feedback in the NSS. Finally, a report by Marsh and Cheng, also published by the Academy, which is concerned with the dimensionality and multilevel structure of the NSS data.

3.1 Surridge: Reports on 2005, 2006 and 2007 data

These reports were commissioned by HEFCE to provide a sector-wide overview of the key findings from the NSS. They are based around two key research questions:

1. What are the key differences between different groups of students in terms of their experiences of HE, as measured by the NSS?
2. What are the variations between institutions once differences in student profiles are taken into account?

In approaching these research questions a range of explanatory variables are used that define different groups within the student body; these are defined according to characteristics of the students, of their courses and of the institutions. The explanatory variables used in these reports are shown in Table 3. In each case the variables are categorical, and therefore different groups must be compared to a fixed 'reference group'; this reference group is indicated in Table 3 in brackets.

In addition to the student, course and institutional characteristics shown in Table 3, the method and timing of the response to the survey were also included to allow for the possibility that students responding by different modes responded in different ways.

Unless explicitly stated otherwise, where findings are discussed below each of the characteristics shown in Table 3 are included in the models being discussed, therefore where an effect is discussed this effect is present after holding all the other factors in the model constant. Models are restricted to full-time students only, except where explicitly stated otherwise. Multilevel models are employed throughout the reports; these are two-level models where students are grouped within institutions. All effects reported as statistically significant within these reports use a 99% confidence level.

Table 3: Explanatory variables used in the reports

Student characteristics	Course characteristics	Institution characteristics
Gender (Male)	Mode of study (Full-time) <ul style="list-style-type: none"> • Sandwich course • Part-time 	Average A-level score on intake (Medium) <ul style="list-style-type: none"> • Low (12-16 points) • High (24-30 points)
Age at commencement of study (18 or under) <ul style="list-style-type: none"> • 19 • 20-21 • 22-30 • 31-40 • 41 and over 	Subject area (Law) <ul style="list-style-type: none"> • Medicine and Dentistry • Subjects Allied to Medicine • Biological Sciences • Veterinary Science • Agriculture and Related • Physical Sciences • Mathematical Science • Computer Science • Engineering and Technology • Architecture, Building and Planning • Social Studies • Business and Administration • Mass Communication • Languages • Historical and Philosophical Studies • Creative Arts and Design • Education • Combined • Two subjects • Three or more subjects 	Size of institution (20,000+) <ul style="list-style-type: none"> • Up to 9,999 • 10,000-14,999 • 15,000-19,999
Ethnicity (White) <ul style="list-style-type: none"> • Black • Asian • Mixed • Other 	Franchised (Not franchised) <ul style="list-style-type: none"> • Mixed or full collaboration 	
Declared disability (None) <ul style="list-style-type: none"> • Dyslexia • Other disability 	Qualification aim (First degree) <ul style="list-style-type: none"> • Other qualification 	
Domicile (UK) <ul style="list-style-type: none"> • Not UK 		
Highest previous qualifications (A-levels) <ul style="list-style-type: none"> • HE credits or higher • Access course • Other 	Research Assessment Exercise (RAE) score of subject area (RAE score 4) <ul style="list-style-type: none"> • RAE score 0-2 • RAE score 3/3a • RAE score 5/5* 	
Term-time accommodation (Other) <ul style="list-style-type: none"> • Living with parents or guardians 		

Key findings: Global trends

Table 4 shows the mean scores for each of the scales and Overall Satisfaction for full-time students, in each case there was a small increase in mean score between 2006 and 2007, in addition to the increases between 2005 and 2006. All of the scales show positive mean scores (above the 'mid-point' of the scale, which is 3).

It is important not to overanalyse the differences between the scales, as each scale is based on differently worded questions and different numbers of items, which may affect the ways in which students respond to the scale. However, it is valid to look at how the scales have changed over time, as each scale is made up of the same items over the three years. The largest increases in mean scores were on the Organisation and Management, Academic Support and Learning Resources scales, both between 2006 and 2007 and over the three years. The Assessment and Feedback scale had one of the smallest increases over time.

Table 4: Mean scores for analytical scales, 2005 to 2007, full-time students

	2005	2006	2007
Teaching and Learning	3.96	3.98	4.00
Assessment and Feedback	3.47	3.49	3.52
Academic Support	3.71	3.76	3.82
Organisation and Management	3.69	3.72	3.79
Learning Resources	3.96	4.02	4.08
Personal Development	3.98	3.99	4.02
Overall	3.96	3.96	4.01

Source: Surridge (2008), Table 2

Key findings: Effects over time for student groups

Tables 5a-c summarise the multilevel models for each of the three years individually. In the tables, a '+' indicates a significant positive effect (that is, the effect listed increases NSS scores relative to the 'reference' category). A '-' indicates a significant negative effect (NSS scores are reduced for this group relative to the appropriate reference group) and 'ns' indicates there is no significant difference between the group and the reference group. In the tables, each cell shows the results for that group over the three years as 2005/2006/2007. Within the table, cells have been highlighted to indicate significant effects across all three years.

The tables clearly highlight the complexity of the effects over the different scales and characteristics. The only characteristic to have the same effect across all scales in all years is method of response, with students responding by telephone more positive across all scales than those who responded online.

Table 5a: Summary of student characteristics, 2005/2006/2007

	Teaching and Learning	Assessment and Feedback	Academic Support	Organisation and Management	Learning Resources	Personal Development	Overall Satisfaction
Gender (Male)							
Female	+/+/+	ns/ns/-	-/-/-	ns/ns/ns	-/-/-	ns/ns/-	+/+ns
Age at commencement of study (18 or under)							
19	ns/ns/ns	ns/ns/ns	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
20-21	+/+/ns	ns/ns/ns	-/-/ns	-/-/-	-/-/-	-/-/-	-/-/-
22-30	+/+/+	+/ns/+	ns/ns/ns	-/-/-	-/-/-	-/-/-	ns/-/-
31-40	+/+/+	+/+/+	ns/ns/ns	-/-/-	-/-/-	+/ns/ns	+/ns/ns
41+	+/+/+	+/+/+	+/ns/ns	ns/ns/ns	-/-/-	ns/ns/ns	+/ns/+
Ethnicity (White)							
Black	-/-/-	+/+/+	-/-/ns	+/+/+	ns/ns/ns	ns/ns/ns	-/ns/ns
Asian	-/-/-	-/ns/ns	-/-/-	ns/ns/ns	-/-/-	-/-/-	-/-/-
Mixed	-/-/-	ns/-/ns	-/-/-	-/ns/-	-/-/-	-/-/-	-/-/-
Other	-/-/-	ns/-/ns	-/-/-	ns/ns/ns	-/-/-	ns/ns/ns	-/ns/ns
Declared Disability (None)							
Dyslexic	-/-/ns	-/-/-	-/ns/-	-/-/-	-/-/-	-/-/-	-/-/-
Other disability	ns/ns/ns	-/ns/ns	ns/ns/ns	-/-/-	-/-/-	ns/-/-	ns/-/-
Highest previous qualifications (A-levels)							
Higher education or higher	-/ns/ns	-/ns/ns	-/ns/ns	-/-/-	ns/-/ns	-/ns/ns	-/-/-
Access	ns/ns/ns	+/+/+	ns/ns/ns	ns/ns/-	ns/ns/ns	+/+/+	+/+/ns
Other	ns/ns/ns	+/ns/ns	ns/ns/ns	ns/ns/ns	ns/ns/+	ns/ns/ns	ns/ns/ns
Term-time accommodation (Other)							
Living at home	+/ns/ns	+/+/+	+/ns/ns	ns/-/ns	+/+/+	ns/ns/ns	+/ns/ns
Domicile (UK)							
Not UK	ns/ns/ns	+/+/+	+/+/+	+/+/+	ns/ns/ns	ns/ns/ns	+/ns/+

Table 5b: Summary of course characteristics, 2005/2006/2007

	Teaching and Learning	Assessment and Feedback	Academic Support	Organisation and Management	Learning Resources	Personal Development	Overall Satisfaction
Qualification aim (First degree)							
Other	-/-/	+ /+ /+	ns /+ /ns	- /ns /-	+ /+ /+	ns /ns /-	ns /+ /ns
Mode of study (Full-time)							
Sandwich	+ /ns /ns	- /- /ns	ns /ns /ns	ns /ns /ns	+ /ns /+	+ /+ /+	+ /ns /+
Franchised (Not franchised)							
Franchised	ns /ns /+	+ /ns /ns	+ /+ /+	- /- /-	- /- /-	ns /ns /+	- /- /-
RAE score (RAE score 4)							
RAE score 0-2	- /- /-	ns /ns /-	ns /ns /ns	- /- /-	ns /ns /-	+ /ns /ns	- /- /-
RAE score 3/3a	ns /ns /-	ns /ns /ns	+ /+ /+	ns /ns /ns	ns /ns /ns	ns /ns /ns	ns /ns /ns
RAE score 5/5*	+ /+ /+	+ /+ /+	ns /ns /ns	+ /+ /+	+ /+ /+	+ /+ /+	+ /+ /+
Subject of study (Law)							
Medicine and Dentistry	+ /ns /ns	- /- /-	- /- /-	- /- /-	+ /+ /+	+ /+ /+	ns /- /-
Subjects Allied to Medicine	ns /ns /+	ns /ns /ns	+ /+ /+	- /- /-	+ /ns /+	+ /+ /+	ns /ns /ns
Biological Science	ns /- /ns	+ /ns /ns	+ /+ /+	+ /+ /+	ns /ns /+	- /- /-	ns /ns /ns
Veterinary Science	ns /ns /ns	- /- /-	ns /ns /ns	- /- /-	ns /ns /ns	ns /ns /ns	- /- /-
Agriculture and Related	- /- /-	ns /ns /ns	+ /ns /+	- /- /-	- /ns /-	- /- /-	- /- /-
Physical Sciences	+ /ns /ns	+ /+ /+	+ /+ /+	ns /ns /ns	+ /+ /+	+ /ns /ns	+ /+ /ns
Mathematical Science	- /- /-	+ /+ /+	+ /+ /+	+ /ns /+	+ /+ /+	- /- /-	ns /ns /ns
Computer Science	- /- /-	- /- /-	+ /ns /ns	- /- /-	+ /+ /+	- /- /-	- /- /-
Engineering and Technology	- /- /-	ns /ns /-	+ /+ ns	- /- /-	+ /+ /+	- /- /-	- /- /-
Architecture, Building and Planning	- /- /-	ns /- /-	ns /ns /ns	- /- /-	ns /ns /-	ns /ns /-	- /- /-
Social Studies	- /- /-	+ /+ /ns	ns /ns /ns	- /- /-	- /- /-	- /- /-	- /- /-
Business and Administration	- /- /-	ns /+ /ns	ns /ns /ns	- /- /-	ns /+ /+	ns /ns /ns	- /- /-
Mass Communication	- /- /-	ns /ns /-	ns /ns /ns	- /- /-	ns /ns /ns	- /- /-	- /- /-
Languages	+ /+ /+	+ /+ /+	+ /+ /+	ns /ns /ns	- /- /-	ns /ns /-	+ /ns /ns
Historical and Philosophical Studies	+ /+ /+	+ /+ /+	+ /+ /+	+ /+ /+	- /- /-	ns /ns /ns	+ /+ /+
Creative Arts and Design	- /- /-	+ /ns /ns	ns /- /-	- /- /-	- /- /-	- /- /-	- /- /-
Education	- /- /-	+ /+ /+	ns /+ /ns	- /- /-	- /- /-	+ /+ /+	- /- /-
Combined	ns /ns /ns	+ /ns /ns	+ /ns /ns	ns /ns /ns	ns /ns /ns	- /- /-	ns /ns /ns
Two subjects	- /- /-	+ /+ /+	+ /ns /ns	- /- /-	- /- /ns	- /- /-	- /- /-
Three or more subjects	- /- /-	+ /+ ns	+ /ns /ns	- /- /-	ns /- /ns	ns /- /-	- /- /-

Table 5c: Summary of institution and method characteristics, 2005/2006/2007

	Teaching and Learning	Assessment and Feedback	Academic Support	Organisation and Management	Learning Resources	Personal Development	Overall Satisfaction
Average A-levels (Medium)							
Lowest	-/-/	ns/ns/ns	ns/-/	-/-/	-/ns/ns	ns/ns/ns	-/-/
Highest	ns/ns/ns	ns/ns/ns	ns/ns/ns	ns/ns/ns	ns/ns/ns	ns/ns/ns	ns/ns/ns
Size of institution (20,000+)							
< 10,000	+/+/+	+/+/+	+/+/+	ns+/ns	ns/ns/ns	+/+/+	+/+/+
10,000-14,999	ns/ns/ns	ns+/+	ns/ns/ns	ns+/+	ns/ns/ns	ns+/+	ns/ns/ns
15,000-19,999	ns/ns/ns	ns+/ns	ns/ns/ns	ns+/+	ns/ns/ns	ns/ns/ns	ns/ns/ns
Response method (Online)							
Post	ns/ns/-	ns/ns/-	-/-/	ns/-/	ns/ns/-	+/+/-	ns/+/-
Telephone	+/+/+	+/+/+	+/+/+	+/+/+	+/+/+	+/+/+	+/+/+
Response time (1-14 days)							
15-28 days	-/-/ns	ns/-/ns	ns/ns/ns	ns/ns/+	ns/ns/ns	ns/-/+	-/-/ns
29-42 days	-/-/	ns/-/ns	ns/ns/ns	ns/ns/ns	ns/ns/ns	-/ns/+	-/-/ns
42 days or longer	-/-/	ns/-/	-/-/	ns/ns/ns	ns/ns/ns	-/ns/+	-/-/

Source: SurrIDGE (2008), Table 3

Key findings: Subject differences

The effects discussed above show there are a number of effects of subject of study that are stable over time. At the JACS1 level (used in the models discussed above), there are 20 subject groupings. Of these the majority showed the same effect, relative to Law, in each of the three years. On the Teaching and Learning scale, 15 of the 20 subjects had the same effect in each year; for Assessment and Feedback this was ten out of 20 (the least stable of all the scales in this regard). For the Academic Support scale, 11 of the 20 groups had the same effect in each year, for Organisation and Management this was 16 of 20, for Learning Resources 13 out of 20, for Personal Development 14 out of 20 and for Overall Satisfaction 16 out of 20.

While it is important to remember that in each case the subject groups are compared with Law, and were the scores for Law to change dramatically in a given year that all the other subject groups would also appear to change, in the three years considered here there is no such change in the scores for Law.

An equivalent exercise can also be undertaken using the JACS2 subject groupings. In this case there are 42 subject groupings to consider. Again, a majority of the subject groups show the same effect in each year. The Teaching and Learning scale shows the greatest stability with 34 of the 42 groups having the same effect in each of the three years. Organisation and Management and Overall Satisfaction have 32 of the 42 groups with the same effect in all three years. The lowest level of stability is found on the Assessment and Feedback scale, but even here 26 out of 42 subject groups have the same effect in all three years. The remaining scales each have 27 out of 42 effects consistent over time.

The high level of stability in subject group effects raise questions of particular interest to educational researchers – why is it the case that, for example, students on courses in Mathematical Science should be more positive about Assessment and Feedback than Law students, but less positive about Teaching and Learning? They also raise the possibility that there may be important lessons that can be learned by looking across disciplines at best practice, while some disciplines may have particular features that are simply not popular with students, it is also possible that some disciplines have evolved practices that other subject areas could learn from.

Key findings: Institutional effects

Institutional effects are measured using the institutional residuals, discussed in Section 2.5 above. Table 6 shows, for each of the analytical scales and Overall Satisfaction, the number of institutions with particular residual profiles. The first column shows how many institutions were performing less well than expected in all three years (had 3 out of 3 negative residuals), the second column those institutions that were performing better than expected in all three years (3 out of 3 positive residuals), the third column shows how many institutions were performing as expected in all the years (non-significant residuals). The table also shows how many institutions were performing better or worse than expected in two of the three years of the survey.

Table 6: Summary of institutional residuals, 2005 to 2007

	3/3 negative residuals	3/3 positive residuals	3/3 non- significant residuals	2/3 negative residuals	2/3 positive residuals	Other combinations
Teaching and Learning	8	9	65	13	6	46
Assessment and Feedback	7	9	55	18	10	48
Academic Support	11	9	64	8	11	44
Organisation and Management	11	9	51	15	12	49
Learning Resources	19	15	49	10	19	35
Personal Development	7	10	85	5	6	34
Overall Satisfaction	10	11	66	8	10	42

Source: SurrIDGE (2008)

As Table 6 clearly shows, the largest group is those institutions that had non-significant residuals in all three years, that is institutions that performed as expected based on their student profile in each year. It should be stressed that the institutions with three negative (or three positive) residuals vary across the different scales. Thus, while there are nine institutions who performed better than expected on the Teaching and Learning, Assessment and Feedback, Academic Support and Organisation and Management scales in each of the three years, these were *not* the same institutions in each case. This has important implications when looking for examples of best practice across the sector, it must be kept in mind that there is variation across the scales, and an institution that is performing better than expected in one area measured by the NSS may not be performing well in another.

These residuals may provide a useful supplement to existing knowledge, within departments and institutions, for those institutions performing less well than expected in two or more years. Institutions are not able to produce this kind of analysis on their own data, and so these institutional residuals provide substantial additional information to institutions on their performance relative to the rest of the sector.

3.2 Fielding *et al.*: Assessment and Feedback, ethnicity and gender

The report by Fielding *et al.* is primarily concerned with differences in degree attainment relating to ethnicity and gender. As part of the wider context of the analysis, the report also contains a detailed analysis of the Assessment and Feedback scale used in the 2006 NSS. In this report we will consider only Section 5 of Fielding *et al.*, which reports on their analysis of the NSS.

The key research aim of Fielding *et al.*'s analysis of the NSS is:

- to explore ethnic and gender differences on the Assessment and Feedback scale, with a particular emphasis on how gender and ethnicity effects are mediated and/or moderated by other factors.

The approach taken by Fielding *et al.* is very similar to that used by SurrIDGE, in that a multilevel modelling framework is used with students and institutions forming the two levels of analysis. This report uses a 95% confidence level, and as this is slightly less stringent than the 99% level used by SurrIDGE we might expect slightly more statistically significant effects to be reported. The analysis in the Fielding report also varies from that of the SurrIDGE reports discussed above in the analysis sample used. The sample used is restricted in a number of ways to make it comparable with other research on ethnic and gender differences in HE. In particular, the sample is restricted in the following ways:

- first degree students only
- valid UCAS tariff scores
- prior entrance qualifications were A-levels or their equivalents
- full-time and full-time sandwich courses only
- UK domicile.

The report uses an analysis strategy that increases the complexity of the multilevel models gradually by adding groups of explanatory variables as well as 'interaction' effects for ethnicity and gender with a range of other variables. Interaction effects estimate the extent to which there are additional effects of belonging to a group with particular combinations of characteristics. For example, there is an effect of being from an Asian ethnic group and an effect of being female, but is there an additional effect (an interaction) of being both Asian and female? To put it another way: is the difference between male and female students the same in different ethnic groups?

Key findings: Gender and Assessment and Feedback

Fielding *et al.* find no statistically significant difference between males and females on the Assessment and Feedback scale. However, effects are more complex when considered in conjunction with other factors using interaction effects. Firstly, when gender is taken together with ethnicity there is an additional effect for females from Indian and Black Caribbean groups: in each case female students from these groups are more positive than their male counterparts. Secondly, there is a significant interaction effect with prior attainment (as measured by the UCAS tariff score). Female students with high UCAS tariff scores are less positive about Assessment and Feedback than are male students with high UCAS tariff scores. Finally, there is an additional effect when combined with age: the gap between males and females is smaller among those aged over 20 than those aged 18 or 19.

A final step in the analysis considers possible interaction effects with subject of study. Here it was found that females were more positive than males in Computer Science, Mass Communications and Creative Arts and Design.

Key findings: Ethnicity and Assessment and Feedback

The modelling strategy employed by Fielding *et al.* begins by including only ethnicity effects in their models and gradually increasing model complexity to assess if additional explanatory variables reduce these ethnicity effects. Twelve¹⁰ ethnic groups are initially used to look at ethnicity effects on the Assessment and Feedback scale. Of these, six groups are significantly less positive than White students: Other White, Indian, Pakistani, Chinese, Other Asian and Other.

The addition of other explanatory variables into the model, including gender, age, disability and subject of study, reduces the size of these ethnicity effects such that four groups remain less positive than White students: Other White, Indian, Pakistani and Other. Further models add UCAS tariff and institutional characteristics, the overall effect being to reduce slightly the impact of ethnicity although leaving the broad pattern unchanged. The general conclusion from this is that there are some ethnic differences relating to the Assessment and Feedback scale, and these are statistically significant for students from Other White, Indian, Pakistani and Other ethnic groups. In each case these groups are less positive than White UK and Irish students.

As with the gender effect discussed above Fielding *et al.* also explored the impact of interaction effects between ethnic group and a range of other factors. This analysis was conducted using a reduced ethnic classification of White, Black, Asian and Other. However, with the exception of subject of study, no statistically significant interaction effects with ethnicity were found (although it should be noted a number of large effects were found, these also had large associated errors meaning that firm conclusions should not be drawn).

The interaction effects with subject of study suggest that Black students were more positive in Sciences or Social Studies (as opposed to Business, History, and Creative Arts and Design), while Asian groups (Indian, Pakistani, Bangladeshi and Chinese) were more positive in Medical Schools.

¹⁰ The groups are White UK and Irish, Other White, Black African, Black Caribbean, Other Black, Indian, Pakistani, Bangladeshi, Chinese, Other Asian, Mixed, Other and Unknown/Refused.

3.3 Marsh and Cheng: Dimensionality and multilevel structure

The report by Marsh and Cheng was commissioned by the Academy and uses data from the 2005 and 2006 NSS. The aims of this research are different from those of the reports by SurrIDGE and Fielding *et al.* discussed above. The specific research aims addressed by this report are:

1. an evaluation of the structure of the NSS questionnaire (in statistical terminology, the dimensionality of the data)
2. to explore the multilevel structure of the data, in particular to look at how much of the variation in NSS scores can be explained at different levels of analysis, that is at the level of the student, the subject area and the institution.

Dimensionality

The report uses the techniques of exploratory and confirmatory factor analysis to assess the structure or dimensionality of the NSS data. The NSS questionnaire groups items into the six scales discussed in Section 2.2 above. In testing the dimensionality of the data there is no assumption that these six areas provide the most appropriate way of grouping the data. Rather, the data are explored to see if there are other ways of clustering the items together that produce stronger relationships between the individual questionnaire items.

The exploratory factor analysis conducted by Marsh and Cheng suggests that the original a priori design of the NSS questionnaire is supported. The analysis suggested that a six-factor solution, which mirrored the six analytical scales described above, was a good fit with the data. However, Marsh and Cheng also demonstrate that a seven-factor solution is consistent with the data. In the seven-factor solution the Assessment and Feedback scale appears to be further subdivided into two factors that relate to fairness of assessment and feedback on assessment.

Marsh and Cheng also use these factor analytic techniques to look at how a factor that measures generalised satisfaction relates to the other scales and to the specific measure of overall satisfaction. The results of these analyses have important implications for the most appropriate way of using the data to measure overall satisfaction. This takes on particular importance in the light of the practice in some league tables of representing overall satisfaction as the unweighted sum of each of the individual NSS questionnaire items.

Marsh and Cheng's analysis suggests that two approaches could reasonably be used to generate a summary of overall satisfaction: i.e. either a summation of the individual items (which should be weighted to take into account the different contribution each makes to overall satisfaction) or the single Overall Satisfaction measure (question 22) could also be reliably used in isolation.

However, while the single item measure is entirely acceptable as a measure of overall satisfaction, it would not be appropriate as feedback to institutions taken in isolation as it is important for institutions to also know which elements that comprise satisfaction are relative strengths and weaknesses.

Multilevel structure

Marsh and Cheng focus exclusively on Overall Satisfaction (as measured by question 22 on the questionnaire) in their analysis of the multilevel structure of the

NSS data. Their modelling strategy is to include three levels in the multilevel models: students, subject groups and institutions. This varies from the approach taken by SurrIDGE (2006, 2007, 2008) where subject of study is treated as an element of the fixed part of the multilevel model to enable comparisons between different subjects. See Section 4.1 below for a more detailed explanation of the differences in the modelling strategies and the interpretation of results.

Using this modelling framework, Marsh and Cheng are able to assess the proportion of the variation in NSS scores that is attributable to different levels within the model, that is what proportion is due to student differences, what proportion is due to subject differences and what proportion is due to institutional differences (these proportions are referred to as variance components). These proportions can be calculated for a range of models that include different explanatory variables. Marsh and Cheng do this for five models. First, they consider a simplified model with two levels: student and institution and no explanatory factors (fixed effects). Second, they add in the intermediate level of subject, again with no explanatory variables. Third, they add to the second model subject differences as explanatory variables. Fourth, they add student characteristics to the three-level model (but exclude subject as a fixed effect) and, finally, in model five both student and subject effects are included in the three-level framework.

The purpose of adding additional explanatory variables to the models is not, in this instance, to examine differences between groups of students, but rather to see if the variance explained by the different levels is reduced when these effects are taken into account. For example, the differences according to subject of study may be in part explicable by differences in the types of student taking courses in specific subjects.

Using these models, Marsh and Cheng estimate that around 2.5% of the variation in NSS scores for Overall Satisfaction is due to differences between institutions. The proportion of variance explained by the subject level is slightly greater, ranging between 3.5% and 7% depending on the exact subject classification used (more detailed classifications explain more of the variance, but also suffer from small sample sizes in some subjects).

The proportion of variance explained at the institutional (university) level is small, as stated above (about 2.5%, once subject differences are controlled for). However, these very small differences are also reliable and stable over time (the correlation between 2005 and 2006 rankings was found by Marsh and Cheng to be 0.86). Marsh and Cheng conclude this creates a dilemma in interpretation:

“The critical question is whether these small differences between universities are sufficiently large to help inform the choices of prospective students ...” (p52)

The intermediate level examined by Marsh and Cheng, subject of study (discipline-within-university), produces complex results from which is it hard to reach definitive conclusions. It is clear from their analysis that a greater proportion of variance is explained by the subject groups within institutions than by differences between institutions. However, it is not clear without further research what the optimal balance is between the number of subjects used and the reliability of results. Due to the potentially small samples in specific subject areas Marsh and Cheng urge caution in comparisons of both different subjects within the same institution or the same subject across different institutions.

Further work

The analyses presented in Marsh and Cheng's report form the first stage of a proposed programme of research that would consider in detail the psychometric reliability of the NSS items and responses. The second stage of this proposed research programme would consider this element in greater detail, looking at how the different NSS components (scales) relate not only to each other but to different student characteristics, as well as how components may be differently related within disciplines and institutions. The third stage of the proposed research would extend the explanatory framework to look at a range of factors that have hitherto not been included in models of NSS outcomes: these would include staff/student ratios, spending on infrastructure and student facilities, completion rates and graduate destinations.

Thus, the work in Marsh and Cheng's report should be viewed as the first step in a wider examination of issues of dimensionality, structure and explanatory exploration and not as definitive findings in these areas.

4. Synthesis: Three years of NSS research

As researchers and other stakeholders increasingly make use of the NSS data to address a range of research questions, it can be confusing as to how far the findings emerging from research are complementary or contradictory, especially where researchers themselves vary in the interpretation of the data.

In this section our aims are threefold: first, to provide a synthesis of research findings (put bluntly, what we know about the student experience as a result of the research so far conducted); second, to highlight areas where research has raised questions but not provided definitive answers; and, third, to address the strengths and limitations of the NSS for understanding the student experience of HE.

4.1 Synthesis

The research reports discussed above each use the NSS data within a multi-variate and multilevel framework; however, the key aims of the research are very different. The research reports by Surridge and Fielding *et al.* focus on differences between student groups, documenting where different groups have different experiences of HE, as measured by the NSS. In contrast, Marsh and Cheng's research focuses on the dimensionality and multilevel structure of the NSS to assess if the NSS is able to provide the kind of guidance to potential students that it was originally envisaged as doing.

Despite these very different perspectives, the findings of the research show a good deal of congruence. Turning first to the research looking at differences between student groups, the reports by Surridge have indicated a high degree of stability in these effects over time. The Surridge and Fielding *et al.* reports also present very similar findings with regard to the Assessment and Feedback scale, despite differences in the samples used for analysis. A general conclusion is that Black students tend to be more positive about Assessment and Feedback than White

students, while students from Asian backgrounds are less positive.¹¹ Both SurrIDGE and Fielding *et al.* found few significant interaction effects with ethnic group, in part due to small sample sizes for analysis. Both reports also highlight the complexity of ethnic effects when ethnicity is disaggregated to 12 categories.

The key area of overlap between the research of SurrIDGE and Marsh and Cheng is the extent to which there are significant differences between institutions and between subject areas. Turning first to the institutional level, both pieces of research suggest that the proportion of the variance explained at the institutional level is small, but nonetheless produces statistically reliable measures of small differences between institutions. The question of the substantive significance of these differences remains open to interpretation. One interpretation is that these differences are too small to be of use to potential students in forming decisions about where to apply for courses. Another is that the small differences should give confidence to students that they will receive a generally good experience of higher education wherever they choose to study and that choice may be guided by other factors, such as disciplinary specialisation. The NSS data cannot discriminate between different interpretations.

The second area of overlap in these pieces of research is in the effect of subject of study on NSS scores. The way in which subject of study is handled in the research varies. SurrIDGE uses a two-level framework (students and institutions) and measures the effect of subject as a fixed effect, while Marsh and Cheng use a three-level framework in which discipline-within-institution is treated as a random effect at an intermediate level between students and institutions. Both of these approaches are statistically appropriate; however, they address slightly different research questions.

The fixed effect approach used by SurrIDGE investigates whether there are differences between subject areas across the sector as a whole, so, for example, whether there are differences between students on courses in Law and students on courses in Mathematics. The random effect approach used by Marsh and Cheng focuses on differences within a single subject area across institutions; for example, whether students on Mathematics courses at one institution are more or less positive than students on Mathematics courses at other institutions.

Despite the different approaches adopted within the research, both SurrIDGE and Marsh and Cheng conclude that there are large differences according to subject of study. SurrIDGE finds that these fixed effects are the largest single effects within the models with differences between, for example, a student on a course in Law and a student on a course in Mathematics being greater than the differences according to gender, age or ethnicity. Marsh and Cheng find that the subject level within their three-level framework accounts for a much larger proportion of the variation in NSS scores than does the institutional level. Both approaches also suggest the need for caution when making comparisons either between different subject groups within an institution or within a subject group across institutions.

4.2 Future research

The potential for further research using the NSS data is high. Marsh and Cheng have set out a programme of research based on developing their framework to include more detailed examinations of the NSS components and their relationship to specific

¹¹ Note: this was true in 2006, the year analysed by Fielding *et al.*, though subsequent analyses suggest the differences may have been reduced in subsequent years.

characteristics, as well as developing the explanatory framework to include a wider range of measures, especially at the institutional/subject level of analysis. There are, however, very many other directions that research using the NSS data could take. Below are some key questions that have been highlighted during the research projects discussed above, but have not yet been pursued.

- The effect of teaching quality on ‘outcomes’

There is untapped potential to link the NSS responses to HE outcomes as measured by degree classifications and graduate destinations. A range of questions are pertinent here as to the impact of different facets of teaching quality on outcomes for students.

- The effect of ‘context’

While there have been some attempts to measure the institutional context in which responses are made, this is an area that could be further developed. Initial explorations suggest that course context (for example, the percentage of ethnic minority students in a given subject area) is more important than institutional context (for example, the percentage of ethnic minority students at a given university). The multilevel structure of the data lends itself particularly well to investigating such effects.

- The impact of student profiles

The models suggest significant differences between different groups of students, and these differences have important policy implications. However, in many cases these significant differences are relatively small in magnitude. Further work could be conducted to assess the impact that different student profiles have on aggregate NSS scores (especially as it is these scores that are made available publicly to aid students in decision making). For example, for a given institution how would scores be affected if the proportion of mature students changed from 5% to 15%?

4.3 The NSS: Strengths and limitations

All research methodologies have strengths and limitations and in this regard the NSS is no different. It should not be viewed as a complete picture of student experiences of HE nor should it be expected to provide answers to all the pertinent questions that may be asked in this area.

The key strengths of a large-scale survey such as the NSS lie in being able to document in detail differences between student groups, particularly small groups, to identify and map trends in student experiences, and in making reliable comparisons between institutions. Analysis of the NSS has, therefore, highlighted key areas of concern that would not be documented in its absence. For example, it has given clear evidence of differences in the experiences of Asian students in some aspects of HE; such a finding could not emerge for individual institutional surveys due to the small numbers involved at any given institution. Similarly, the NSS has allowed us to

map out subject differences across the sector, potentially highlighting ways in which different subject areas may learn from each other.

The NSS is particularly useful for highlighting areas of concern, in other words it is good at identifying the 'where' in student experience: where there are differences in experiences as rated by students. It is not, however, designed to answer the 'why' of student experience: why some groups of students have more positive experiences than others. It cannot, for example, tell us *why* Asian students have less positive experiences of some aspects of their courses or *why* there are large and stable differences among subject areas.

It is important to remember these limitations and not to expect the NSS to provide all the answers. At an institutional or departmental level, it provides data that must be supplemented by on-the-ground knowledge if to be used effectively. At the level of the sector as a whole, it points to issues that are of concern for policy but can only be fully grasped via other modes of analysis, perhaps via work with small groups of students.

5. The NSS: What does it mean for us?

5.1 Potential students and the NSS

The NSS data have added to the wealth of information already available to potential students about the institutions to which they are considering applying. The extent to which this information is used by potential students and their advisers is monitored by Unistats. What contribution to this process can be made by the research reported here? First, it should give some reassurance that the differences between institutions, albeit that they are very small, are reliable and not artefactual. Second, the research may give confidence that the vast majority of students at HE institutions in the UK have very positive experiences with regard to the measures of teaching quality included on the NSS.

However, the research has also highlighted the need for some words of caution in using the data as publicly provided. First, some of the units (courses and/or subjects) with published data have relatively small numbers of respondents and therefore relatively high associated statistical error terms. Second, the data as published for use by potential students do not take into account differences between student groups. While this is unlikely to be a major biasing factor for most students who are likely to be comparing similar subjects and types of institution, it could be misleading were a student making comparisons, for example, between studying Biological Sciences or History.

5.2 Institutions and the NSS

The NSS provides a valuable source of data to individual institutions; however, the dataset as received by institutions is relatively limited and does not lend itself to the kinds of analysis discussed here. The research discussed in this report should be seen as providing an additional resource to institutions: a contextualisation of their own results in the light of results elsewhere in the sector.

Three elements of this research are key for institutions. First, the need to contextualise NSS scores on the basis of the types of students at an institution. The research has shown that different groups of students have different experiences of HE and institutions should aim to be sensitive to the experiences of the types of students that make up its student body. This is also important in making comparisons between different parts of an institution, comparing the raw scores of the Mathematical Sciences department with those of the History department may be misleading where the composition of those departments is different and where the subject areas vary substantially across the sector as a whole.

Secondly, the research provides a contextualised measure of institutional performance via the institutional residuals. These residuals are an important supplement to on-the-ground knowledge about different aspects of teaching quality at an institution. They provide a measure of performance not directly available to institutions as they show performance relative to the rest of the HE sector. In other words, while an institution's own NSS results may be improving, it is possible that it is still performing worse than expected relative to other institutions, if the other institutions are improving at a faster rate.

Finally, the research suggests that the differences between institutions are stable and reliable. It also shows that these differences are small, either as measured by the number of institutions that have statistically significant residual terms or as measured by the proportion of variation explained by differences between institutions. The vast majority of students rate their higher education experiences positively in the NSS, and the vast majority of institutions are not statistically different from each other in this regard.

5.3 The HE sector and the NSS

The research discussed in this report takes the HE sector as a whole as its primary research interest, and therefore many of the key findings have greatest resonance at this level. Key findings based on the experiences of specific groups of students raise questions that must be addressed at the sector-wide level and that in many cases require further research; a particularly striking example is the experiences of some aspects of teaching quality for Asian students within the HE sector as a whole.

The three pieces of research all highlight the importance of subject-based differences in experiences of HE. This raises the question of whether there are important lessons that may be learned by subject areas looking across the sector at other subject areas, as well as the more usual comparisons of good practice within subject areas.

6. Published research and further information

Published reports:

Fielding, A., Charlton, C., Kounali, D and Leckie, G. (2008) *Degree attainment, ethnicity and gender: Interactions and the modification of effects. A quantitative analysis*. Available from:
http://www.heacademy.ac.uk/assets/York/documents/ourwork/research/EDA_Quantitative_Report_March08.pdf.

Marsh, H. and Cheng, J. (2008) *National Student Survey of Teaching in UK Universities: Dimensionality, Multilevel Structure, and Differentiation at the level of University and Discipline: Preliminary Results*. Available from:
<http://www.heacademy.ac.uk/ourwork/research/surveys/nss>.

Surridge, P. (2006) *The National Student Survey 2005: Summary report*
— *The National Student Survey 2005: Response and survey methodology*
— *The National Student Survey 2005: Findings (main report)*
All available from http://www.hefce.ac.uk/pubs/rereports/2006/rd22_06/.

Surridge, P. (2007) *The National Student Survey 2006 - Summary report*
— *The National Student Survey 2006 - Full report*
Available from http://www.hefce.ac.uk/pubs/rereports/2007/rd14_07/.

Surridge, P. (2008) *The National Student Survey 2005-2007: Findings and trends*
— *Summary report: Non-traditional students*
— *Summary report: Institutional variation and change over time*
— *Summary report: Part-time students*
All available from http://www.hefce.ac.uk/pubs/rereports/2008/rd12_08/.

Other sources of information:

HEFCE Learning and Teaching: National Student Survey

<http://www.hefce.ac.uk/learning/nss/>

Higher Education Academy: National Student Survey pages

<http://www.heacademy.ac.uk/ourwork/research/surveys/nss>

IPSOS Mori National Student Survey information

<http://www.ipsos-mori.com/nss>

Unistats

<http://www.unistats.com>

Appendix: Multilevel models

When data have an inherent structure then many of the assumptions of standard statistical techniques may be violated leading to misleading or erroneous conclusions from these models. Multilevel models allow us explicitly to model the structure of the data. These models can be thought of as having three basic components – the fixed part of the model, the random part of the model and the model residuals.

Fixed part

The fixed part of a multilevel model produces estimates of the effects of particular characteristics on NSS scores. These estimates are interpreted in exactly the same way as those of a traditional single-level regression model. In general, there are three things to look at for each estimate: the size of the estimate, the direction of the effect and the statistical significance of the effect. The size of the estimate tells us how big a difference to NSS scores the particular characteristic makes, in general this is the effect of a one-point change in the characteristic on the NSS scale being measured.

However, this is further complicated as in many of the models the characteristics are categorical variables, that is they are made up of a number of different sub-groups rather than a numerical scale. For example, when considering the effect of ethnic group we have a number of categories of ethnicity (White, Asian, Black, Mixed and Other) that cannot be properly represented numerically. For these types of data we estimate an effect for each of the sub-groups relative to one group that remains fixed (known as the reference group). For example, when looking at the effect of ethnic group we have an estimate for the effect of being from Black, Asian, Mixed and Other ethnic groups, which in each case is relative to the effect of being from a White ethnic group (our reference category). In these instances, the estimates are more simply interpreted as the effect as a result of belonging to the group being estimated rather than the reference category, and it is especially important when interpreting these effects to be aware of what the reference category is.

The direction of an effect is measured by its mathematical sign: positive numbers indicate that the effect increases scores on the given NSS scale, while negative numbers indicate the effect reduces scores on the NSS scale.

Finally, each estimate is checked for statistical significance; this tells us how confident we can be that the effect found in our sample of respondents is actually present in the student population as a whole, as all estimates are subject to an element of random variation according to the particular subset of students who responded to the survey. In general, NSS results are very reliable as we are dealing with very large sample sizes, this means that even effects that are very small (that is they only change NSS scores by a small margin) may still be statistically significant.

Random part

The random part of a multilevel model allows us to see how much of the variation in scores is due to differences in individual students and how much is due to variation among a higher level unit (for example, institutions). It is this part of the model that specifically takes into account the structure of the data and the commonalities between individual units grouped according to the higher levels (for example, the commonalities among students at the same institution).

Institutional residuals

All statistical models have 'residuals', which represent the difference between the data as collected and the outcome predicted by the model. In a multilevel model these residuals apply to each level of the model, therefore there are a set of residuals that show how much the student score is different from the score predicted for a given individual student and there are also residuals that show how much the institutional outcomes vary from those predicted by the model. These 'institutional residuals' are a measure of institutional performance after we have taken into account all the factors in our models.

For example, we may find that students on courses in Medicine are less satisfied with an aspect of their course than students in other subject areas. This would lead us to expect that institutions with Medical Schools would, all other things being equal, achieve lower scores than institutions without Medical Schools. Raw NSS scores do not take these kinds of factors into account; however, the residuals from the models do take into account variations like these. Therefore, institutional residuals may be used as a measure of institutional performance that has taken into account the types of students at the institution.

As with coefficients in the fixed part of the model, these institutional residuals must be assessed for statistical significance. This is done by creating a confidence interval around each of the institutional residuals. A confidence interval is a range of values within which we expect the value for the population as a whole (as distinct from the sample we have analysed) to lie. Confidence intervals can be created for different confidence (sometimes called significance) levels. The most commonly used confidence levels in social science research are 95% and 99%. These levels mean that if we drew 100 different samples from the data and calculated the residual, in 95 out of 100 (or 99 out of a 100 in the case of the 99% confidence level) samples the residual value would lie within the range calculated.

This is illustrated in Figure 1 below. Each point in Figure 1 represents an institutional residual shown with a bar to illustrate the 99% confidence interval (were a 95% confidence interval used these bars would be shorter). As can be seen from the figure, the majority of institutions have confidence intervals that cross the zero line and are therefore not statistically significant from the scores we would expect for that institution. The circles highlight a small group of institutions at each end of the chart; those in the bottom left are institutions whose confidence interval lies entirely below zero. This indicates that these institutions had scores that were significantly lower than those we would expect from their student, course and institutional characteristics. While the circle at the top right highlights a group of institutions whose confidence interval lies entirely above the zero line, this group of institutions had scores that were significantly higher than we would expect from their student, course and institutional characteristics.

In other words these two groups of institutions (as highlighted by the circles in Figure 1) are statistically different from the 'average' institution after we have taken into account their student profile. Those in the bottom left have scores that are significantly worse, while those at the top right have scores that are significantly better. The majority of institutions are not statistically different from the 'average' institution once differences in student characteristics are taken into account.

Figure 1: Institutional residuals + 99% confidence interval, Teaching and Learning scale, 2006

