The perception of academic staff in traditional universities towards the National Student Survey: views on its role as a tool for enhancement

Adam W A Child
M.A. in Education (by Research)

Department of Education
University of York

September 2011

Abstract

The National Student Survey (NSS) has been a part of the higher education landscape since 2005. Since it was first mooted the NSS has been a controversial topic within academia, with many people expressing concern about the robustness of the survey and many others seeing it as an important way for students to express views about their programmes.

This thesis explores the perceptions of academic staff towards the NSS and seeks to establish the ways in which the NSS is currently used within higher education specifically for the purposes of enhancing learning and teaching. The research questions of this study relate to these issues as well as exploring the differences between disciplinary areas. A wide-ranging literature review was firstly undertaken to set the political scene and determine the extent of the previous work in this area. This in turn led to the development of a mixed methods approach, with both qualitative and quantitative data gathered from over three hundred academic staff via an online questionnaire. The analysis chapters feature both descriptive statistics and a regression analysis in order to respond to the research questions.

The conclusions of this study make the argument that the NSS is not necessarily seen as suitable for concurrently performing the three main functions it is seen by policy makers as achieving. Therefore further consideration needs to be given to the way people engage with the data produced by the survey. There were not any major differences between academic staff from different disciplines. However this could be more because of the seemingly generic nature of the NSS, which in turn contributed to a general scepticism about the survey. The implications from this study are explored at several levels: departmental, institutional and national.

Contents

List of tables and figures

Acknowledgements

I would like to take this opportunity to thank the many people who have contributed to the development of this dissertation. Top of the list is my supervisor, Dr Paul Wakeling, whose sound guidance has aided me greatly. Equally important is the moral support of my partner, Emma, who has coped with many a weekend spent alone whilst I analysed data in our spare bedroom.

I would also like to thank those members of the academic community who found the time to complete the questionnaire. Without their contribution this dissertation would not have been possible.

Adam Child

1. Introduction

This dissertation explores the perceptions of higher education academics towards the National Student Survey (NSS). The NSS has been a feature of the United Kingdom higher education landscape since it was first run in 2005. Since then it has become increasingly high profile, often featuring as a major part of the commonly used league tables (CHERI et al, 2008).

1.1. The National Student Survey: a bone of contention

In 2010 a major review of the National Student Survey took place, the findings of which suggest that the future role of the NSS is three-fold: to provide a number of metrics informing prospective students about aspects of the course they may wish to study (Oakleigh Consulting and Staffordshire University, 2010); to provide information for use in quality assurance processes and to support enhancement activities within institutions (Centre for Higher Education Studies, 2010). The authors of the report *Enhancing and Developing the National Student Survey* remark that,

> the NSS should continue to support all three of these dimensions – student choice, quality assurance and quality enhancement, the last of these being directly related to the student learning experience**.** We found striking the emphasis that *institutional managers* placed on the way the NSS findings allowed them to identify potential problems in the student experience, and to act on them quickly (Centre for Higher Education Studies, 2010, p3). (emphasis added).

It appears the case that the decisions being taken about the future of this national survey are primarily with the needs of the institutional managers in mind. Nowhere in this particular review of the NSS were the views of individual academics assessed to see if they value, or even use, the NSS for the purposes of enhancement. This dissertation therefore is an attempt to explore this issue, and redress the balance between the needs

of the senior management within institutions and the lay academics who are ultimately responsible for delivering higher education. When the NSS has been described as "generally accepted across the [higher education] sector" (CHES, 2010, p9) this has not been based on robust information about the perceptions of academics towards this survey. It is this gap that this dissertation attempts to address in order to contribute to policy discussions about the uses of this national survey.

There has been a great deal of media coverage surrounding the NSS since its inception. In the early years of the survey some student unions attempted to boycott the survey as they saw it as intrusive and over simplistic (Cambridge University Students' Union, 2010). In addition, and perhaps more worryingly for those who support the survey, are the range of criticisms from members of the academic community. One particularly strong attack came from Harvey (2008) who described the NSS as "Shallow, costly, widely manipulated and methodologically worthless". This view was supported by other stories that had emerged in the higher education press about the manipulation of the survey at Kingston University, where two members of staff were accused of telling their students to produce high scores in order to maintain the prestige of their course (Mostrous, 2008). More recently a number of humanities academics at the University of Brighton have criticised the NSS, despite the fact their institution actually achieved a top ranking for that subject area, describing it as a "statistically risible exercise in neoliberal populism". (Attwood, 2010). These concerns seem to suggest an underlying myriad of perspectives on the survey, as opposed to the feelings of general acceptance reported in the works commissioned as part of the Teaching Quality Information review.

1.2. The research question

With the above debate in mind, the following research question is proposed for this study with a number of supplementary issues:

What are the perceptions of academics towards the National Student Survey and its use as a tool for enhancement?

- Is the NSS perceived by academics to be a reliable indicator of teaching quality?
- Do academics use the results of the NSS for enhancement purposes and what are their motivations for doing so?
- How do academics usually use the data (if at all)?
- Are there any differences between academics of different disciplinary backgrounds?

Underlying these research questions is the broader question of whether the perceptions of academic staff have an influence on the acceptance of strategies which intend to provide evidence to justify teaching and learning interventions. Does the acceptance (or otherwise) of an intervention lead to a greater engagement with it? It is expected that this research may contribute to the debates around evidence informed practice in higher education as well as the specific debate about the NSS.

My interest in this topic stems from the time I spent working at the Higher Education Academy where I developed a general interest in the use of student surveys to inform the direction of enhancement activities. Throughout this time there have been a multitude of individuals based within institutions who have expressed concern on behalf of themselves or a general constituency of their colleagues, that the NSS is being used in an inappropriate way, or that too much credence is being placed on the results. This led me to compare this general feeling with policy level discussions, showing an apparent mismatch. This mismatch seemed to provide an interesting issue to investigate further.

1.3. The research strategy

The work I have conducted in a professional capacity at the institutional and disciplinary level has contributed to the development of a research

strategy that seeks to investigate the differences between these two paradigms of belonging. It has been argued that the academic can both adapt to the institutional context when developing enhancement activities (Gibbs, 2000) or, alternatively respond only to the issues arising from their own discipline thus resulting in generic initiatives from government or senior management being less well received (Becher, 1994).

The first stage of the research strategy was therefore to review the literature exploring three specific topics relating directly to the research questions. One area for investigation was the development of student evaluations of teaching and the associated research exploring the perceptions of academics towards these evaluations. Much of this literature comes from outside of the United Kingdom. The second element of the literature review concerned the development of the National Student Survey itself, which includes detail of other models that could have been adopted, as well as the concerns of the sector when the current tool was in its infancy. The third aspect of this initial investigation was the literature around disciplinary differences to provide a research based infrastructure for any differences that became apparent in the later stages of the research.

The next part of the strategy was to gather intelligence about the issues pertaining to the NSS within an institutional context. As outlined in the methodology chapter this was largely done through a series of interviews with academics to explore their perceptions and see what the underlying issues were. This was an important way of determining the types of questions to ask during the next part of the research strategy. The third part of the strategy was to survey the academic community, although at the early stages of this research project it was not apparent what form of survey would be most appropriate. For example, a small number of qualitative interviews could have been conducted instead of the questionnaire approach that was later favoured. The final part of the research strategy was to ensure some form of qualitative data was available to inform the analysis. As this research is exploring perceptions

it was deemed of crucial importance that the “voices” of academics on this topic were heard, alongside any statistical analysis conducted. As will be highlighted below, the data was collected through a single questionnaire. However the differences in the types of data being collected called for different forms of analysis and as such this research uses a mixed-methods approach.

1.4. Chapter commentary

This chapter has introduced the general area in which this study resides and discussed some of the debates surrounding the NSS and its use as a tool for enhancement. This has shown a mixed picture requiring some additional exploration. The chapter has also introduced the research strategy that will be built upon further in the methods chapter.

Chapter two explores the literature available in three specific areas. The first of these is the background on student evaluations of teaching. There has been a huge amount of work done in this area, and much of this has shown the evaluations to be useful both for rating teaching and for assisting in the formulation of interventions to improve classroom provision. The second part of the review looks at the background history of the NSS and establishes the reasons for the development of the survey and the ways in which the survey has been analysed to establish its validity and reliability. Crucially this chapter also explores the current policy level discussions about the NSS, with the most recently commissioned work endorsing the use of the NSS as a source of public information and as a means of enhancing higher education provision. The third part of the review shows the differences established through empirical research between higher education disciplines. This section shows the potential for disciplinary differences to be an interesting part of this research as there are notable differences between subject groups.

Chapter three provides a detailed account of the way in which the research strategy has been developed. The first part of the chapter

explains why the particular research method has been chosen. The chapter then leads into an account of how the questions for the questionnaire were developed. This was a multi-stage process. Firstly there was a series of interviews with academic staff within universities to help frame the questions for the questionnaire. The questionnaire was then piloted with a small number of staff within one of the randomly selected institutions. The last part of chapter three describes the way the main questionnaire was distributed to the collected sample.

Chapters four and five provide the detailed analysis of these data with a view to answering the main research questions. Chapter four provides an overview of the quantitative data and explores the qualitative data provided in response to the two open comment questions. A number of analyses are conducted on the quantitative data, for example a reliability analysis and an ordinal regression of the core seventeen questionnaire items. Chapter five looks at the differences between the three disciplines chosen for this study as well as the differences between institutions of different types and parts of the UK.

The final chapter develops the conclusions of the study and answers the research questions specifically. Chapter six also discusses the implications of the study at several different levels of higher education. This chapter also evaluates some of the issues with this study and proposes some potential directions for further study.

2. Literature Review

There is a wide range of literature relevant to the current study. This literature review has been separated into three parts. The first section explores the research around the use of student evaluations and reveals the potential they have to be used for enhancement purposes but also shows a mixed picture of the ways staff perceive their usefulness. The second section focuses on the development of the National Student Survey, its origins and the reasons behind its creation. The third section delves into disciplinary perspectives around learning and teaching. The general picture emerging from this part of the literature has implications for the current study as it reveals the differing views of non-cognate disciplines towards teaching and learning implying the possibility of differences at a disciplinary level towards the use of the NSS.

2.1. Student evaluations of teaching

There is a significant body of evidence showing that student surveys are used in a widespread fashion across higher education, in many countries. The use of these surveys for the improvement of teaching depends on several contextual factors and on the perceived validity of the specific survey instrument.

Student surveys have been used at the module level within Higher Education Institutions since the 1920s to provide information to academics about their personal performance and the quality of their provision (Flood Page, 1974). The use of these surveys is particularly prevalent in the United States. Work by Murray (1997) found that by the mid 1970s the majority of institutions in the United States were using surveys of this type. Murray assumed the reason for this was the widespread body of evidence showing these surveys to be generally reliable; related to other objective measures of teaching; correlated with assessments by fellow staff members and only mildly affected by factors

such as class sizes (Murray, 1997). This assessment matches with the previous work by Marsh (1987) which showed that in general student surveys do correlate favourably with the quality of teaching and learning. Although there is no single study proving this correlation beyond doubt, Marsh looked at a number of ways in which teaching quality was assessed for example self-evaluation, peer-evaluation and external observation, finding these to correlate with the ratings provided by students. Marsh also explored a number of statistical techniques used to confirm the validity of student evaluations. So there is evidence available showing survey tools to be potentially useful sources of information for improving student learning and informing teachers about their practice.

There is a wide range of literature exploring the ways in which information arising from student surveys has been used to improve student learning. Cohen found as early as 1980 that on the whole feedback had a “modest yet significant” effect on improving instruction. Cohen’s work brought together the findings of a multitude of studies. Cohen’s other major finding was the greater improvement found in those studies where feedback was augmented by a consultancy process with a third party (Cohen, 1980). This finding was echoed in later work by Marsh and Roche (1993). Ballantyne et al (2000) agreed that this augmentation process was of paramount importance. The authors felt that evidence was lacking about the effectiveness of student evaluations in isolation and that surveys could be used to identify specific staff development needs. The usefulness of student feedback came from the engagement students had with other faculties enabling them to identify weaknesses more readily than the staff themselves (Ballantyne et al, 2000).

As Murray (1997) suggests in his North American study, there are logical reasons why student evaluations should lead to an improvement of teaching. These include: the motivation to achieve tenure through good results; the added motivation to seek the help of a consultant and the general help feedback on an activity can provide (Murray, 1997). By the 1990s there were moves to develop surveys that were specifically

designed to facilitate improvement of provision and teaching. Cashin and Downey (1992) developed a tool for summative assessment of teaching and proposed a longer form of this survey that could be used for diagnostic purposes. There was potential for the summative judgement to spur on improvement activities (Cashin and Downey, 1992). Another example was the work done at the University of Hertfordshire where a student survey was developed based on the Course Experience Questionnaire within the School of Engineering. This survey was designed specifically to lead to enhancement activities. The results of the survey were reported up to committees within the University who then developed an action plan which pivoted around further qualitative investigation and discussion with students. There were two issues arising from this work, the first being that the pace of change was slow and difficult to attribute to changes made after the surveys. Secondly it was easy to take up issues with the survey instrument itself and lose the developmental aspect of the process (Gregory et al, 1995). This is something worth investigating further in the context of using the National Student Survey as an enhancement tool; does the perception of the tool effectively block the route to any worthwhile enhancement activities?

Kember et al (2002) questioned the widespread value of student questionnaires on the basis that in some departments they were rarely discussed and individual teachers were alienated by the amounts of data. In the institution studied by the authors, the instrument was standardised and imposed upon departments which could have led to an impression that the survey lacked credibility (Kember et al, 2002). Yorke (1995) also picked up this theme by suggesting that a single instrument could not be used in all contexts for managing quality but could be used at an institutional level as a broad performance indicator (Yorke, 1995). This raises questions for the current study. If the National Student Survey is seen as an imposed instrument lacking applicability to a departmental context, does this affect the way it is used for enhancement?

The solution proposed by Yorke was to develop a range of interventions which can be used to provide points of cross reference (Yorke, 1995). One suggested addition was concept mapping (Saroyan and Amundsen, 2001). The importance of being able to allow for multiple contexts was put into perspective by d'Appolonia and Abrami (1997). They suggested that there are now many instructional contexts, including internships, interactive seminars and computer assisted instruction. If the definitions of instructional effectiveness are based on the products and processes of instruction they do not necessarily generalise across these other contexts (d'Appolonia and Abrami, 1997). Using multiple measures of teaching removes the need for one reliable, specific measure. This could allow student surveys to be interpreted as a measure of student perceptions rather than teaching quality per se. Even if a student was giving wholly subjective views on teaching, it still provides information to the teacher about the way the student perceives it. The teacher needs to be able to interpret this information (Falk and Dow, 1971).

Each evaluation of teaching, using surveys or other methods, has as its heart an underlying concept of teaching that influences both the teachers and the students during the evaluation process. Kolditch and Dean (1999) suggested two paradigms: "Transmission" and "Engaged-critical". They argued that the student survey they were observing seemed to assume the "transmission" model, effectively discounting the other paradigm. They found that this could actually be to the detriment of teaching quality as good scores would be sought by changing behaviour towards a teacher-centred learning style (Kolditch and Dean, 1999). A large scale Australian study found that students have very different educational upbringings influencing their views of the ways they were being taught. The differences in disciplines also had an impact. This meant that establishing the causes of a poor score was very difficult (Timpson and Desley, 1997). Although this might be seen to be less relevant to a survey like the NSS that is conducted at the end of a programme of study, work described below (Flint et al 2009) indicated that students completing the

NSS were thinking across the whole programme whilst completing it, meaning educational upbringing might still be relevant.

The detailed differences between disciplinary perspectives are outlined below but in the literature about student evaluations these differences are featured. Nasser and Fresko (2002) created a questionnaire to allow faculty staff to provide their views on student evaluations. They found a significant difference between teachers of different disciplines but they did not offer any detailed reasons for this difference. Richardson (2005) made a tentative statement about the disciplinary differences when he suggested that open ended questions to gather qualitative information could be useful for programmes in the humanities where students are often sceptical about the value of using quantitative information for understanding the world (Richardson, 2005). By extension this could provide a practical difference between teachers of different disciplines also.

There is a range of articles exploring the perceptions of academics towards student surveys. An early polemic against using student surveys for evaluation of teaching was provided by Kerlinger (1971) who suggested that surveys of this type could actually lead to an erosion of the relationship between teacher and student. In terms of improving teaching, Kerlinger believed that student evaluations could actually be to its detriment (Kerlinger, 1971). Not too much emphasis should be placed on this article as it amounts to no more than the view of one individual. The general concerns of teachers were summarised by Flood Page (1974) who found that worries included the idea that students are too inexperienced to rate their teachers and they would in turn rate popular teachers as good teachers (Flood Page, 1974). On this latter concern, Dent and Nicholas (1980) found in a survey of students and staff that the majority of faculty staff believed that student surveys could influence a teacher to seek the favour of their students. Interestingly however the students believed this was not the case (Dent and Nicholas, 1980). Schmelkin et al (1997) went one step further to argue that the anecdotal

evidence point towards a “widespread hostility” towards student evaluations.

One of the key themes of the literature is the perceived validity of the student evaluation tools. Where student evaluations are seen by faculty as valid, there seems to be a positive relationship with their overall view of the tools. An example of this was a survey of faculty revealing this relationship as well as what the authors described as a “self-interested rationalism”, in other words those who did well in student evaluations are more likely to approve of them (Nasser and Fresko, 2002). A possible example of this is clear from the views of the academic staff in this study towards league tables (see section 4.2). This could explain the interest of some authors in the link between positive results on student evaluations and reward and recognition mechanisms within universities. One survey of 25 departments suggested that the use of the data for improvement of provision was linked to incentives for improving results (Kember et al, 2002). Marsh and Roche (1993) saw these incentives as possibly leading to a desire for the improvement of teaching quality.

The current study is looking at the views of academics towards the National Student Survey and its role as a tool for enhancement. There is a lack of literature exploring the views of academics towards a national level survey. Some tangentially related literature can provide some idea about the general view of teachers towards these surveys and their potential to enhance teaching. A study using the Course Experience Questionnaire to facilitate activities around learning and teaching found it was very easy to take issue with the questions of the survey and lose focus on the actual content of the results (Gregory et al, 1995). An article focusing on semi-structured interviews revealed that the majority of lecturers felt student surveys had caused them to think about their teaching although individually very few had actually made specific changes in their provision (Moore and Kuol, 2005). The implication is that the perceptions of the tool seem to have an impact on the level of its use

and this can be hypothesised to be the case with the National Student Survey.

2.2. The development of the National Student Survey

The National Student Survey was first administered in 2005 and has been analysed in detail to show its validity and reliability. To date there is very little literature exploring its use as a tool for the improvement of student learning and less still that studies the views of academics towards the survey in a systematic way.

In 2000, the Higher Education Funding Council for England (HEFCE) proposed to replace the extensive review mechanisms currently in place with the publication of key data on quality matters to help prospective students make informed judgements on where to study, and thus help discharge the accountability function of a sector in receipt of large amounts of public money (Richardson et al, 2007). In response to this, a task group chaired by Sir Ron Cooke wrote the 2002 paper *Information on quality and standards in higher education* (HEFCE, 2002). In this report three principles for accountability in higher education were recommended, these were:

- Meeting the need to provide information to the public;
- The responsibility of the institutions to use robust procedures and publish key information;
- Having systems that are relatively light touch. (HEFCE, 2002)

The development of the National Student Survey can be seen as part of this broader change of the quality assurance system.

The first UK move towards a national survey of students originated in a 2003 report to HEFCE. The report took the view that there were two components that had to be balanced, accountability and improvement. It

was suggested that a national survey could provide information to be used for accountability purposes. The improvement activities were intended to be informed by internal processes, as these could be tailored to a departmental or institutional context (CHERI et al, 2003). In other words, the national survey of students was never intended to be used as an enhancement tool. The feeling of the staff who were interviewed as part of this research indicated a feeling that a national survey would add little to internal feedback mechanisms already in place and the results would be too general to provide useful information (CHERI et al, 2003). This begs the question as to whether this perception before the development of the NSS is actually borne out in reality and makes the present study all the more relevant.

The report by CHERI et al reviewed the current survey models in use at the time to capture data related to student experiences. The report focused primarily on two models, the student satisfaction approach (Harvey, 1997, 2001) and the Course Experience Questionnaire (CEQ) (Ramsden, 1991). It was concluded that a new model, based on the CEQ would be the way forward. The student satisfaction approach was heavily criticised due to the lack of robust evidence about its validity and the more fundamental issue with having student satisfaction as a goal for higher education (CHERI et al, 2003). The two approaches are fundamentally different. The student satisfaction approach has as its central purpose the improvement of quality within institutional contexts (Harvey 1997). The CEQ was developed initially as a performance indicator to justify governmental expenditure on higher education, enhancement of teaching was seen as a “positive side-effect” (Ramsden, 1991). Interestingly both Ramsden (1991) and Harvey (1997) see the student as a consumer of higher education and this is used as justification for the necessity of their respective approaches. Wiers-Jenssen et al (2002) would later agree during their assessment of student satisfaction as a concept that students have a right to evaluate their own evaluators (Wiers-Jenssen et al, 2002). Another important difference is the level at which the two surveys work. Ramsden’s CEQ surveys

students at the programme level meaning no teacher gets individualised results. Harvey's approach works primarily at a more abstract level, institution wide, but he sees the need to augment this with module level feedback (Harvey, 2001). This module by module student evaluation would impact on the individual teacher.

When considering the development of the NSS it is important to note the underlying principles behind its sister survey, the CEQ. As Hanbury notes in her comparison of national surveys, "Being based on the CEQ, the theory base of the NSS is the same as for the CEQ, i.e., it emphasises the importance of students' perceptions of their learning context and the impact of this upon their learning outcomes" (Hanbury, 2007, p10). It is the positivity of these perceptions which the CEQ seeks to provide information on. The rationale behind the CEQ when it was developed in the early 1990s was as a performance indicator. The survey was designed to provide information that could be used across institutions as module level evaluations were too unsystematic and varied to provide meaningful information on that scale (Ramsden, 1991). Ramsden was able to counter the common concerns about student surveys by suggesting that the CEQ only posed questions about the areas of students' experience that they are qualified to comment on (Ramsden, 1991). A further analysis was carried out a number of years later by Wilson et al (1997). In this analysis the CEQ was used alongside the Approaches to Studying Inventory to determine relationships between the types of responses provided to the CEQ and the type of learning shown by the individual students. This study found there to be a correlation between positive scores on the CEQ and a deep approach to learning. There was a negative correlation with a surface approach (Wilson et al, 1997). This suggested the appropriateness of using the CEQ as a proxy for the quality of the student experience. This was a major motivation for using the CEQ as the model from which to build an equivalent survey in the UK.

As early as 1994, researchers were looking at the potential of the CEQ for the UK context. Richardson conducted a survey using the CEQ questions and collated 95 responses. Analysis of the results found a series of correlated first order factors that were “dominated” by a second order factor. This second order factor was related to the students’ experiences of their course and in one sense this factor could be used as a measure of quality. Richardson also suggested that the assessment scale required further work as the strength of this correlation with the second order factor was less strong (Richardson, 1994). The sample size of this study was very small, but the indication was clear – the CEQ had some relevance within the UK higher education context.

Small scale studies conducted within medical education provided some interesting findings concerning the application of the CEQ. One study by Broomfield and Bligh (1998) surveyed 180 medical students using the short form of the CEQ. They concluded that the CEQ was an “appropriate instrument for course evaluation” (Broomfield and Bligh, 1998). However, question marks were raised during a subsequent study undertaken by Lyon and Hendry (2002) as they sought to evaluate the usefulness of the CEQ in assessing the quality of a problem based learning medical course. The CEQ was used as justification, combined with general faculty concerns, for the changes of a medical course to incorporate a problem based element (Lyon and Hendry, 2002). In general the authors found that the changes they had made to the course were favoured by the students with two notable exceptions – the “Clear Goals and Standards” and the “Appropriate Workload” scales. The authors concluded that the CEQ was created at a time when courses were taught in a different way, with a high level of teacher regulation. This led to the generation of clear goals with students knowing what is expected of them. In a problem based learning environment, students are more anxious about whether they are covering the correct material and are left to trust that they will learn what they need to know (Lyon and Hendry, 2002). This raised a legitimate question about the validity of the survey tool for the modern university system.

The first run of the NSS was in 2005 and surveyed approximately 280,000 students in their final year as an undergraduate. The report following the pilot one year earlier suggested that there was still room for improvement in the survey and it had to be shortened in length. It also had to be determined whether the survey captured the essential dimensions of teaching quality (HEFCE, 2004). What is most interesting about this report are the warnings it offered about the inability of the overall satisfaction question to be used as a publishable result and the need to avoid using the NSS as a way to compare institutions across the whole sector without taking account of the individual institutional contexts (HEFCE, 2004). However, both inter-institutional comparisons and the publication of NSS results are now occurring in the form of league tables.

The impact of the league tables published by newspapers such as *The Times* and the *Guardian* was investigated as part of a broader study commissioned by HEFCE and published in 2008. In three of these league tables, those from *The Times*, *The Sunday Times* and *The Guardian*, the NSS scores were weighted more heavily than any other factor. The findings of the report suggested that this has had an impact on the profile of the survey within institutions. Senior institutional managers were found to be increasingly interested in improving their NSS scores and this had led to a top down model for developing enhancement activities (CHERI et al, 2008). One research intensive university in the study felt the use of the NSS in the league tables had increased the profile of the student experience within their institution; another university explained that in direct response to the NSS scores they had improved their facilities for students and course organisation (CHERI et al, 2008). What is not clear from the study is the view of the individual academic based in a department towards this development. The report by CHERI et al makes it clear that the institutions care a great deal about the scores. Whether the staff in faculties and departments share this enthusiasm is a question for the present study.

The most detailed analysis of the NSS datasets was conducted by Marsh and Cheng (2008). The aims of their study were to test the structure of the NSS instrument and to determine how much of the variance was attributable to background statistics such as discipline (Marsh and Cheng, 2008). There were several key findings from this study. One interesting finding was the suggestion that the overall satisfaction question (Question 22) could actually be appropriately used as a summary score. Marsh and Cheng also found that some subject areas, such as History and Philosophical Studies had a higher average score than other areas. This leaves us with the question as to whether the teaching is more effective in this subject across the board, or whether there is something inherent in the subject leading students to rate it more positively (Marsh and Cheng, 2008). It is this difference between global subject areas which led Marsh and Cheng to conclude that meaningful comparisons could only be made between units of different disciplines when they were within the same institutional context. Discipline units of the same subject could be compared across universities (Marsh and Cheng, 2008). It was concluded by Marsh and Cheng and later by Surridge that comparisons using the NSS data had to be exercised with caution. (Marsh and Cheng, 2008; Surridge, 2009). The work of Yorke (2009) built upon the previous work of Surridge (2008) in demonstrating the reliability of the survey instrument. Yorke tested the survey by changing the order of the questions and the order of the Likert scales to test for order effects or acquiescence bias. He did this by redistributing the survey randomly in lectures with different question orders to see if there was an effect on the way the surveys were completed. He found little effect and therefore suggested that this should be a “reassurance” to those who designed both the NSS and the CEQ (Yorke, 2009).

There is not much literature exploring the potential of the NSS for developing enhancement opportunities. Peer reviewed material is especially slight. There was concern expressed by Williams and Kane (2008) that activity of this nature was rarely taking place and the concern was more about appearing further up the league tables. The authors

recommend the use of action plans and the incorporation of the student voice, potentially using Harvey's student satisfaction model (Harvey, 1997, 2001).

One article by Flint et al (2009) describes the use of the NSS alongside a series of student focus groups. This approach seemed to be a positive experience as it allowed the more detailed unpicking of the survey results within the institutional context. Two findings emerged from this study; the first showed students to consider their experience at the course level rather than the individual module level. The second was that students rarely understood the impact of their feedback on the development of undergraduate provision, largely because they had not been adequately informed of the changes made (Flint et al, 2009). The idea that students view their experience at a programme level could lend extra validity to the NSS, as it measures the perceptions of students at that level. However, this leads us to question the ownership of NSS-led enhancement by individual academics and the potential for centralisation of enhancement activities into the hands of a few enthusiasts. When considering the perceptions of academics towards the NSS, the manifestation of these perceptions also has to be considered. How do lecturers respond to the survey, what do they actually do?

2.3. Disciplinary differences

The literature around disciplinary culture within higher education is well established and generally accepted. What is less well developed is the idea that this should impact on views towards specific teaching and learning tools and mechanisms although this has been looked at in a limited way. There is a widespread assumption within the academic community that the average academic feels affinity to their disciplinary area in a fashion that outshines their affinity to their institution or to a broader notion of academia. In 2000 the *Improving Student Learning Symposium* focused on this specific issue (Rust, 2000). In the same year, the Learning and Teaching Support Network (LTSN) was established to

provide support to academics through twenty-four subject centres. The eleven year existence of this network was recognition of the disciplinary element of learning and teaching enhancement. The role of the NSS in contributing to this disciplinary level enhancement activity is one that is gathering momentum as the NSS becomes more prominent nationally.

The lines drawn between disciplines or groups of disciplines are often cultural. The first significant attempt to articulate the cultural differences between subject groups was made by Biglan (1973). In this article, which has been widely cited since, Biglan studied two colleges in the United States asking academics to group subjects together. He then assigned specific characteristics to those subjects. From this activity, Biglan was able to develop a typology of subjects with common characteristics. This was based on two spectra: hard-soft subjects and pure-applied subjects. Hard subjects have a tendency to use numerical data, whilst soft subjects often emphasised qualitative information. Pure subjects are generally more theoretical, applied subjects are more grounded in reality. (Biglan, 1973)

Although Biglan's study only surveyed a couple of institutions, his typology seems to have been accepted by later writers in this area. Becher (1994) took Biglan's work and developed it further by adding detail to the stereotypes around each discipline area, including common criticisms of particular subject types. As an example, soft-pure subjects are sometimes seen as not being relevant to the outside world thus leading to the lack of outside funding for these subjects. Becher also cites the importance of the wider communities that have an influence on subject areas, including professional bodies. The idiosyncratic nature of each subject area leads Becher to express concern at the nature of both generic performance indicators and non-discipline specific faculty development programmes,

> Faculty development programmes, for instance, tend to lose credibility with their potential clients because of their discipline independent

> approach… It is difficult to see how faculty development can go beyond the most elementary level without a clear recognition that disciplinary cultures impose their own particular pattern in teaching as in other activities. (Becher, 1994, p158)

The NSS is one of many generic performance indicators currently operating in higher education. If Becher's views are reflected across the current community of academics, it could result in a lack of enthusiasm towards using the NSS as an enhancement tool.

An interesting take on the differences between specific disciplines was taken by Braxton who assessed the nature of disciplines in turn in an attempt to establish which subjects, if any, had a natural tendency to improve teaching and learning in university education (Braxton, 1995). The central pillar of his work is an assumption that those in soft disciplines are more interested in general character development and thinking skills, in comparison with the hard disciplines who focus more on facts and concepts. The emphasis on character development in the soft disciplines in turn creates an affinity within the teachers of those subjects to develop their teaching more, to naturally promote deep learning. The student-centred approaches which are generally recommended in higher education are more likely to occur within these affinity disciplines (Braxton, 1995). Graham Gibbs continued this theme by suggesting that disciplines and departments have their own cultures and that these are easy to pick up. These cultures are likely to have implications for some elements of the chosen teaching approaches, for example the level of democracy within the department. Gibbs' contribution to this debate is important as he suggested that these cultures are not necessarily "hard wired" and are often born out of tradition and convention. They can be changed to suit a pedagogic requirement (Gibbs, 2000). The question may be to what extent lecturers are willing to challenge conventions for the development of their teaching. This has a significant implication for the current study. Do the academics within soft disciplines take on board

the results of the NSS more readily than those based within hard subject areas?

Analysis of the specific teaching styles generally used in certain subject areas has been conducted by a number of researchers. Neumann's Australian study found soft subjects to have a tendency to use established techniques such as programme review to improve their provision and in turn the student evaluations these subjects received were generally more positive than the scores received by the hard subjects. Neumann concluded his work by warning about the notion of applying common teaching and learning techniques, including performance indicators, slavishly. A claim that one teaching method is better than another has first to take account of disciplinary differences (Neumann, 2001). Neumann et al (2002) continued on this track, with an article that developed generalisations about the ways hard and soft subjects were taught. Hard-pure subjects tend to be in larger groups and support teaching with handouts, projector slides and the like. On the other hand, soft-pure subjects often use seminars and the occasional high-profile lecture; tutorials are also used, which allow an individual's own perspective to be aired (Neumann et al, 2002). This contrasts slightly with earlier findings from a survey conducted in the early 1990's in Norway finding that although content is often decided through disciplinary norms, the methods used are often decided within the institutional context (Smeby, 1996). Neumann et al (2002) also has something crucial to say in light of the current study, namely that student evaluations are specifically one of the areas where the assumption that all disciplines are similar can cause an issue. Explanations for the consistently high ratings of soft disciplines have been offered but it is seen as unlikely to be due simply to their teaching being coincidentally better in these subjects and is more likely to be a result of complex cultural and epistemological differences. This quality procedure therefore fails in practice (Neumann et al, 2002) particularly if the overall summative assessment does not take account of these factors. The argument is that one is not comparing like with like when comparing disciplines and therefore the comparisons are

unfair. Only when reasonable comparisons are made can an evaluative tool be used for comparative purposes. With one of the aims of the NSS being explicitly around the issue of accountability in UK higher education (CHERI, 2003; Richardson et al, 2007) this is a significant issue requiring further investigation.

Hatvia and Birembaum (2000) explored the preferred learning styles of students through a survey of 175 students in Education and Engineering schools. They found a similarity between the preferred teaching styles of the two cohorts, with both groups favouring the “providing” instructors over the “self-regulation” instructors. This surprised the authors as the “self-regulation” form of instruction is generally favoured by educational developers. Their conclusion was that the students adapted to the learning style presented to them, adopting a surface or deep approach as a result. This however, did not affect their actual preferences (Hatvia and Birenbaum, 2000). This conclusion links well with the work of Entwistle and Tait (1995) and Ramsden and Entwistle (1981). Entwistle and Tait found that students prefer to be taught in a style which is familiar to them. Subjects requiring rote memorisation often lead students to take a surface approach to learning and this is prevalent in the science subjects. How students respond to their learning environment is linked to their perception of the environment (Entwistle and Tait, 1995). The importance of perceptions of the learning environment on student learning style was explored by Ramsden and Entwistle (1981). They found a relationship, through a widespread survey of 171 departments in 54 universities across six disciplines between teaching method, approach to learning and perceptions of student experience versus self-reported progress. The authors of this study felt it was now the responsibility of individual departments to create a learning environment to foster the deeper approaches to learning (Ramsden and Entwistle, 1981). This article was important for another reason as it started to demonstrate the link between student perceptions of learning experience and quality of learning. This is one of the underlying principles behind the development of the NSS and its predecessor surveys across the world.

The implications of the perceived and real differences between disciplines require further investigation if we are to establish how academics use the NSS as an enhancement tool. Do the disciplinary cultures have any effect on the way academics use the NSS? Are some disciplines more likely to use the NSS for this purpose than others?

There are several interrelating factors at play that could influence the findings of the current study. The cultural differences between departments of different subject areas may be significant, as may be the broader institutional contexts that colour the perceptions of individual academics. The opinions of individuals towards the survey may be born out of these contributing contextual aspects but what is not so clear is which factors are most influential and how these shape the way the survey is used for improving teaching within universities. It is these questions the present study seeks to probe.

3. Methodology

3.1. Choice of research method

The literature currently available at the time of writing suggests that a broad based study of a large sample of academics exploring their views towards the NSS would be the best method to increase understanding of the ways this national scale survey is used. The research questions outlined above refer to academic staff working within UK higher education. This is a broad group and it would prove difficult to provide generalisable answers to the research questions without consulting the views of a large number of staff from different institutions and backgrounds. With this consideration in mind a questionnaire for completion by academics was developed. Questionnaires can often be misunderstood as being simply a way of collecting quantitative data. They can more accurately be regarded as collecting systematic data (De Vaus, 2002) and because of the requirement of the present study to collect data from a large number of people, a questionnaire is the most appropriate method. Marsh (1982) saw the method of using a questionnaire as being often open to criticism for not allowing the development of the full internal nature of the group being studied and for causing an atomisation of complex social structures. However she argued that these criticisms are often levelled at poorly designed questionnaires and concerns can be addressed by ensuring that the questionnaire asks questions at the appropriate unit of analysis (Marsh, 1982). Of course in addition to this the questionnaire has to be appropriately pretested and piloted.

It was decided at an early stage to develop an online questionnaire as this research method has several advantages over mail and telephone surveys including time and cost (Dillman et al, 2009). The use of online surveys, even in today's technological society is not without its critics. The appropriateness of an online survey is often determined by the level of information technology literacy within the population being surveyed.

Dillman et al (2009) suggested that people's lack of comfort with Internet technologies may also be a factor in deciding whether or not to take part in an Internet survey. However, in this particular case this was not expected to be a significant problem as the nature of UK higher education means the vast majority of academic staff will have ready access to computers and the Internet on a daily basis. The substantial increase in the use of IT to deliver higher education courses (Enyon, 2005) suggests academic staff are well-equipped to respond to an online survey. In order to ensure that further data could be collected in the event of a low response rate a question was added asking respondents to leave their email addresses to facilitate follow up interviews (see Appendix 2).

The detail required to provide reasonable responses to the research questions meant both quantitative and qualitative questions would be needed. Independent methods of analysis were to be used on the two types of data and this could therefore be considered to be a form of mixed-methods research despite the fact both forms of data were collected via the same instrument. Mixed methods research is defined as being the collection of both numerical and word-based data (Greene et al, 1989). Mixed methods approaches are gaining favour as a way of investigating social phenomena. Johnson and Onwuegbuzie (2004) suggested that approaches categorised as mixed methods fit together qualitative and quantitative data in a workable solution. They argued mixed methods approaches would also provide a superior product to studies using a single method. In the current study the range of question types within the questionnaire is an effort to provide internal triangulation; achieving what Creswell (2003) saw as a cancellation of the biases of a single method. The collection of qualitative and quantitative data is concurrent, but this is an accepted strategy for mixed methods research (Creswell, 2003). In this particular study however, the mixed methods approach taken was unable to cancel the sampling bias, which was caused by the sampling method employed during this research (see section 3.4).

3.2. Pre-survey scoping

In order to inform the development of a series of questions for the pilot survey there were two main stages of pretesting with members of the academic community. The first stage was to hold a series of informal (i.e. not recorded), unstructured discussions with academic staff with whom the author already had an established relationship. Seven of these face to face discussions were held with the general topic of conversation being the use of the National Student Survey (NSS) within the context of the individual concerned. This stage of pretesting revealed some interesting findings leading to the development of specific questions for the pilot questionnaire. For example a recurring theme was the “departmental culture” loosely defined as the manner in which the management of a department view the concept of students passing judgement on the performance of the department. This pervading culture could be a source of encouragement for those wishing to use NSS data, or an issue if the views of students are not accepted as being valid on this topic. There was also a very clear message emerging about the relatively top-down nature of interventions informed by the NSS.

The first stage of pretesting informed the development of a series of questions designed to probe the issues outlined above as well as the key research questions forming the basis of this study. Once these draft questions had been developed, the second stage of pretesting was to show these questions to 12 colleagues from within the higher education sector. This was to check for a consistent understanding of the items and ensure there were no questions which people would be unwilling or unable to answer. There were three changes made to the questionnaire as a result of this pre-test. For example, the draft questions had used as one of the Likert statements, “The NSS is a valid survey”. This statement was seen as rather vague due to the differing conceptions of survey validity. The intention of this statement was to unpick whether or not academics felt the NSS measured the quality of a learning experience. It was decided that there were other more appropriate statements which

probed the same issue in a less ambiguous way. In total 12 people reviewed the proposed survey questions; some of these individuals had experience of survey design themselves, others were specialists in other areas of academia. The approach of using “expert review” of a survey tool as part of a pilot phase was suggested by CHERI et al (2003) in their report to HEFCE on the topic of collecting student feedback. It has also been suggested as good practice when developing survey tools (Newman and McNeil, 1998) and as something that could potentially improve or change the findings from a study (Davis, 1992).

The questions for the pilot questionnaire largely consisted of Likert scale items followed by a series of demographic questions designed to provide background information about the respondent’s university, department and job title. Likert scales are a tool for developing composite measures of a concept and used in combination can provide a full picture. It is the use of Likert items in combination that allows them to be most useful and less misleading (De Vaus, 2002). Likert scales can be designed in a number of different configurations, with an odd or even number of points as the researcher deems appropriate. Preston and Colman (2000) found scales with fewer than four points to be less reliable and those of ten points or more of having less test-retest reliability. The authors concluded that scales have to be designed with methodological and practical concerns in mind with shorter scales having the benefit of being relatively quick and easy to use. Taking account of these considerations it was decided to use a five point Likert scale for the current study because of the time pressures on staff being asked to complete it. Cox (1980) also confirmed that five points seemed adequate for items which are subject-centred in approach (as is the current study). The people who read the questions as part of the pre-test stage felt comfortable with the use of a five point Likert scales for asking questions on this topic.

One interesting comment was made about the use of negatively worded statements. In the draft questions there were a number of negatively worded statements requiring the understanding of a double negative. A

number of people suggested a change in the wording of these statements to the affirmative. The literature on research methodologies reveals a lack of consensus on the use of negatively worded statements. An early study from the U.S. concluded that only affirmatively worded statements should be used to avoid confusion (Wembridge and Means, 1918). A later study by Wason and Jones (1963) found that the use of the word “not” in a questionnaire has a prohibitive effect on the person completing the instrument and those responders generally translate negative statements into affirmative statements during their cognitive response to the question. This second point is also indicated by the extra time the respondents took to answer the negatively worded items (Wason and Jones, 1963). A different study found higher levels of random error and lower levels of reliability in negative survey items (Muircheartaigh, Krosnik and Helic, 2000). A study in the realm of higher education showed some statistically significant differences between the responses to positively and negatively worded items and argued that responses may be coloured by factors outside of the concepts being measured (Weems et al, 2003). A paper by Paulhaus (1991) suggested a nuanced solution where both acquiescence bias and the use of negatively worded statements are avoided by the adding of conceptual opposites as affirmative statements (Paulhaus, 1991). This approach was taken with the pilot survey and the negatively worded statements were amended to reflect this.

3.3. The pilot questionnaire

Before the main questionnaire was distributed, a pilot was planned to ensure the survey worked technically as well as check that the final questionnaire was viable in terms of response rate, completion rates and quality of information collected. To develop a sample for the pilot one institution was chosen at random from the institutions selected for the main sample (see 3.4). This university’s website was then viewed in order to gather details from the departments of Chemistry, Psychology and English. These disciplines were chosen as cognate subjects to the main disciplines chosen as part of the main study. This was done out of a

desire not to “use up” any of the contacts that had been collected for the main sample. The sample size for the pilot was 92.

The pilot run of the questionnaire was launched on the 4th August 2010 and the “Pro” version of Survey Monkey was used to collect the responses. Survey Monkey is a popular online survey tool enabling users to collect large sets of data and has facilities for the use of a wide variety of question styles, including Likert scale questions; open response questions and multiple select questions. The responses were collected in a way that did not allow the answers to be traced back specifically to the individual respondent, increasing the importance of the demographic questions towards the end of the questionnaire. This raised the potential issue of having more than one response from an individual; however the survey tool has a function to block multiple responses from the same IP address, which would work to minimise this problem. The questionnaire was distributed using an email mail merge function available through Microsoft Outlook. This allowed each email to be tailored using the background information gathered about the sample from the departmental websites including first name, surname and title. This was seen as an advantage for improving response rates. A reminder email was sent on the 15th August 2010.

A problem which instantly manifested itself was the blocking of emails containing the words “Survey Monkey” by the institutional spam filter, thus preventing any of the emails from being received. This was managed by changing the survey link to one that would not be picked up by institutional filters. This was a useful technical point arising out of the pilot. Another anticipated problem was the accuracy of the data about academics on departmental websites. With there being a time lag between the collection of the email addresses and the distribution of the questionnaire it was thought that there would be a number of staff changes. During this pilot only one email “bounced back” due to a member of staff having left the university, showing that perhaps this may not be as significant a problem as first thought. The likelihood of this

being an issue for the main questionnaire was increased as a new academic year had started between the collection of the staff details and the distribution of the main round of emails.

The first email generated 15 responses and the reminder led to a further 11 responses. This total of 26 represented a response rate of 28%. An analysis of the response rates of early email surveys suggested that the response rates for surveys of that type were falling steadily and early surveys had been more successful in generating good response rates. There were two factors suggested as being able to combat this trend. The first is to ensure that the survey is salient to the sample and the second is to avoid unsolicited surveys (Sheehan, 2001). On this first point, the pretesting stage was designed specifically to maximise the relevance of the questions to the population. On the issue of unsolicited questionnaires, Sheehan (2001) does indicate that for a number of studies, some form of unsolicited contact is unavoidable. It was decided before the pilot stage to attempt to personalise the initial contact as much as possible to reduce the perception of the email being unsolicited. In order to prevent frustration on the part of the sample members and reduce administrative burden, the link to the questionnaire was included as part of the first contact (and subsequent follow ups). Ethical protocols were observed during all contact with academic staff. The email sent to staff provided a basic introduction to the study as well as the name and direct contact details of the researcher. In addition, respondents were assured that the responses would be anonymous and they were invited to request a copy of the final thesis if they wished. A copy of the email is available in Appendix 1. The pilot stage of the study showed the potential of the questionnaire to gather sufficient numbers of respondents which in turn would allow for some robust conclusions to be drawn from the dataset.

Some tentative initial analyses were conducted on the pilot questionnaire results using PASW version 18, although the sample from the pilot was obviously very small. The main part of the questionnaire, consisting of 17

Likert scale questions was tested for internal reliability using Cronbach's alpha. Cronbach's alpha provides a co-efficient of equivalence for a test which uses a series of items (Cronbach, 1951). The figure from the test provides some idea as to the amount of variance provided by one underlying factor, i.e. the concept being tested. The initial reading on this test was 0.453 which is below the commonly accepted threshold of 0.6 indicating a reliable scale. This however was expected due to the nature of some of the items, which were affirmatively worded but were actually intentionally testing the reverse concept i.e. antagonism towards the NSS. The scores in these items were reversed so that their numerical value was in the same direction as the rest of the items. The Cronbach's alpha increased to 0.631 showing the satisfactory internal reliability of these Likert scale questions. The removal of the item "My institution could use the data more effectively than it currently does" would have increased this coefficient to 0.707 This naturally led on to an exploratory factor analysis of the pilot data to see if this would be a useful technique to employ during the final analysis. An exploratory principal components analysis suggested five factors with an Eigenvalue greater than one. These factors were extracted with a varimax rotation. This revealed there to be a number of items that loaded on to more than one factor, but there were also nine of the items that only loaded on to one factor. This suggested one underlying concept was being measured by the questionnaire tool but also that there were unlikely to be easily identifiable underlying factors loading onto this.

An experimental cross tabulation of the pilot data using subject area and the answers to the core 17 items revealed no statistically significant differences between the responses on the basis of subject area. The majority of the Chi-square significance levels were >0.100, when <0.05 is required for the differences to be considered statistically significant. This was almost undoubtedly a symptom of the small sample size, especially within subject areas (ns ranged from 13 to 4 respectively). At the pilot stage this served as additional motivation for maximising the sample size and ensuring there was good representation from each of the chosen

subject areas for the main questionnaire. This would be essential to minimise the potential for non-response bias and ensure that the dataset was as useful as possible. The same issue would have likely been the case when seeking to make comparisons between institutions, although the pilot did not allow more detailed exploration of this as it focused on one university.

There were two qualitative questions as part of the pilot questionnaire. One invited respondents to comment on the ways the department or faculty used the results from the NSS. The other question asked them to comment on the way they used the NSS data in an individual capacity. These questions were designed to flesh out the detail about how academics use the data thus answering one of the core research questions for this study. The level of response to these questions was encouraging, with 80% of respondents answering the question relating to departmental use of the NSS and 85% responding to the question about how they use the NSS as an individual. The level of detail provided by the respondents was surprising, with several people providing paragraph-length responses. If this was repeated across the main questionnaire it would provide a very rich source of qualitative data and negate the need to do any form of interviews with individual academics to supplement the data; although this remained an option as academics were asked to leave their email address. The added advantage of gathering qualitative data in this way is the ease with which cases can be categorised by subject or institution in a programme such as NVivo, thus enabling the form of analysis demanded by the original research questions.

3.4. Sampling for the main questionnaire

The sample for the full questionnaire administration was developed in a systematic way. There were several options available for the development of the sample. It would have been possible to survey all academic staff within a small number of institutions (or possibly just one institution). It would have also been possible to survey academic staff

within one discipline across a large number of institutions. However, neither of these extremes would have allowed for the comparison of views towards the NSS between both disciplinary and institutional types. Therefore it was decided to randomly sample institutions and then take the details of academic staff who were based within three disciplinary areas. The three disciplinary areas chosen were History, Physics and Education. These choices were made because they reflected personal interest and because each of these subject areas sat within a different part of Biglan's typology of subjects in higher education (Biglan, 1973). It was the time limited nature of this study which meant that the compromise between disciplinary scope and number of sampled institutions had to be reached and it was understood that this would leave the "Hard-applied" category of Biglan's typology uncovered. However, despite this the sample was deemed sufficient to allow an initial analysis of the dataset by discipline area, although it is recognised that further research in this area should widen the disciplinary scope to enable more generalised conclusions to be drawn.

Technically this sampling frame is regarded as a multi-stage cluster sample with institutions as the sampling units and the departments acting as secondary units. As Barnett (1991) suggests, often forms of cluster sampling are selected for reasons of pragmatism. However in this particular case this form of sampling fits well with the original research questions by allowing comparisons to be drawn between particular lists of respondents within each primary and secondary unit of analysis.

The very different mission of research-intensive and research-led institutions in comparison with teaching-led institutions means that the conclusions of the present study cannot be slavishly applied to the whole of a diverse higher education sector. In order for this wider perspective to be explored, further research would need to be undertaken. Each institution was chosen in turn using a random number generator readily available on the Internet[1] with numbers being assigned to each institution corresponding to their position in the 2010 *Sunday Times* league table.

This table was chosen simply because it contained the largest number of institutions.[2] When an institution was selected the university website was searched for the departments or faculties in the three selected subject areas. If an institution did not have a department or faculty in one of the subject areas the university was rejected for the purposes of this study as this would not allow within-institution comparisons. This selection of the three subject areas in this study did inadvertently create an extreme bias towards universities designated before 1992, as these were more likely to have the three required departments. The choice of disciplines effectively excluded Pre-92 institutions from the sample. This was an unintentional bias in the selection of the sample that had to be accounted for at the analysis stages. It could not be assumed for example that the results for this set of data would be applicable to the rest of the sector, including Post-92 institutions and Further Education Colleges.

If an institution had the required departments their departmental websites were then visited to gather the publicly available information about individual members of academic staff, including their email addresses to enable the electronic distribution of the survey. One difficulty with this method of collecting information about academic staff was the differing classification of staff within each institutional context. For example, the job title “Research Fellow” can indicate a member of staff who is research-only, or could show a member of staff with a teaching portfolio. In each case the information on the website was interpreted to establish whether they were likely to be a teacher or not.

In total an initial sample of 1308 academics was collated from 12 institutions across the United Kingdom. Every nation of the UK was represented by at least one university. Although this was not originally an intention of the method of sampling, it meant comparisons between the nations were more likely to be possible and this analysis was later carried out (see section 5.2).

Although the sample was collected systematically and with the research questions of the study in mind, there were a number of biases within it that needed to be identified before conclusions could be drawn from the available data. The choice of disciplines was a non-random decision which in turn meant that some members of the higher education population were less likely to be selected than others either because they did not work in that subject area or because they worked in an institution which did not teach in those subject areas. Sampling bias is common is sociological research as often the sample is selected in a non-random fashion (Winship and Mare, 1992). This sampling bias manifested itself in the selection of Pre-92 institutions, meaning that the staff surveyed were all from this specific type of institution. This has an effect on the external validity of the results, that is, the ability to infer conclusions from the data about those who did not feature as part of the sample. It would not be possible for example to assume the findings of this study apply to academic staff within Post-92 universities or those teaching subjects outside of the trio defined here. Although this is not intended to discount the usefulness of undertaking the present study, the limitations of the sample have to be understood in order to make realistic conclusions from the data.

3.5. The main distribution of the questionnaire

The encouraging results from the pilot stage of the questionnaire and the smooth technical running of the survey meant that no changes were deemed necessary, either to the method by which the survey was distributed or to the items forming the questionnaire. The first wave of emails asking the full sample to participate in the questionnaire was sent on the 17th October 2010. This generated 208 responses before the reminder email was sent on the 1st November. The survey was closed on the 8th November and 324 responses had been collected by this point. As anticipated during the pilot stage, the lag of time between the collection of information about individual members of staff and the distribution of the survey did mean there was a slightly higher number of emails bouncing

back, indicating that a member of staff had changed institution. There were also a number of people who sent an email back indicating that they believed they would not be able to provide a useful response due to having either recently retired, or not having an undergraduate teaching load. These people were taken out of the sample, leaving a sample size of 1250. The response rate for this survey was therefore 25.9%. Again, this was roughly in line with expectations following the pilot stage. The number of responses and the amount of qualitative data produced meant that the option of conducting follow-up interviews with some of the respondents was not required.

The responses were then imported into PASW 18 and coded as appropriate to allow the data analysis to take place. The first part of the analysis was to explore any differences between the original sample and the actual group who responded to establish whether there were clusters within the sample left unrepresented and in turn show where there may be a bias due to non-response. Tables 1-4 below provide a breakdown of the demographic details of the respondents.

Table 1 shows the proportion of the sample from each of the randomly selected universities and compares this with the percentage of responses provided by each institution. The largest difference is for University 3, where there is an 8.1% difference between the size of the target sample and the percentage of respondents. This is a large difference and this will have to be taken into account when the analysis relating to this institution is compared to the overall descriptive statistics. The rest of the differences are in the region of 2/3% and although these may make a difference, these are unlikely to be statistically significant. Two of the universities have fewer than 10 responses against them, and although these were the two with the smallest numbers in the questionnaire sample it does mean that when conducting institution level analysis one has to be careful about generalising the views of staff at those institutions on the basis of a small sub-sample. One way to combat this would be to group institutions in some way (e.g. type, region etc) as a way of making

broader conclusions about institutions of a certain ilk. This is the form of analysis explored in chapter five.

Table 1: Frequency table showing the number of respondents from each sampled university and university group

University no.	Group	Percentage in sample	Response frequency	Response rate (%)	Percentage of responses
University 1	N/A	4.4	7	12.7	2.2
University 2	Russell	10.5	26	19.8	8.0
University 3	Russell	18.6	34	25.8	10.5
University 4	1994	9.8	30	24.6	9.3
University 5	Russell	8.1	27	26.7	8.3
University 6	N/A	5.8	21	28.8	6.5
University 7	N/A	3.8	8	16.7	2.5
University 8	Russell	9.8	34	27.6	10.5
University 9	Russell	9.4	30	25.4	9.3
University 10	Russell	7.7	30	31.3	9.3
University 11	1994	4.8	16	26.7	4.9
University 12	N/A	7.3	22	24.2	6.8
Sub-total		100	285	N/A	88.0
Non responses		N/A	39	N/A	12.0
Total		100	324	25.9	100

Table 2 shows the differences in the discipline specialism between the target sample and the actual respondents. Due to the entry of some "others" during the questionnaire and some missing values, the comparison has to be between the questionnaire target sample and the percentage of responses against one of the three desired subject areas: Education, History and Physics. Two hundred and ninety-three of the questionnaire responses were against one of these subject areas. History is slightly overrepresented in the sample of responses; the other two disciplines are slightly underrepresented. This will only cause an issue for the analysis at the macro-level if there is a statistically significant difference between subject areas on any of the substantive items of interest. As mentioned earlier in the literature review, Braxton (1995) found that some disciplines have a culture showing more affinity with issues relating to the quality of higher education teaching and learning. It

was originally hypothesised prior to the pilot that this may lead to more Education academics completing the survey due to the general salience of the topic. This does not appear to be borne out in the response rates, although this does not preclude it making a difference when the analysis is conducted at a disciplinary level.

Table 2: Frequency table showing the number of respondents from each discipline

Subject area	Percentage in sample	Frequency	Response Rate (%)	Percentage of responses
Education	36.4	96	21.1	29.6
History	33.5	117	27.9	36.1
Physics	30.1	80	21.3	24.7
Other	0	19	N/A	5.9
Sub-total		312		96.3
Non responses	N/A	12		3.7
Total	100	324	25.9	100

Table 3 provides detail of the job titles provided by respondents during the questionnaire and makes a simple comparison with the job titles gathered from publicly available departmental websites during the sampling process. In the survey tool, this question was asked using an open comment box, which led to a much wider variety of job titles in the questionnaire responses when compared with the original sample. When this is combined with the introduction of missing values it leaves a very complex picture. However, without any detailed calculations it is possible to review the job titles of those who responded and make a qualitative comparison. This would seem to suggest that the proportions of people with each job role compare reasonably well with those of the original target sample. There are no large anomalies; an original concern prior to the pilot stage was that only those staff with a major teaching component to their role would feel well placed to respond. It appears that the questionnaire was of relevance to a broad range of staff at different levels within their departments.

Table 3: Frequency and percentage of academics of each job title

Job title	Percentage in sample	Frequency	Percentage of responses
Lecturer	33.2	97	29.9
Senior Lecturer	15.0	47	14.5
Reader	6.2	28	8.6
Professor	25.8	81	25.0
Director	1.4	11	3.4
Head of Department[3]	N/A	6	1.9
Research Fellow	3.9	7	2.2
Teaching Fellow	5.9	12	3.7
Other[4]	8.6	6	1.9
Retired/Unemployed	0.0	2	0.6
Response total		297	91.7
Non responses		27	8.3
Total	100	324	100

Table 4: Gender breakdown of respondents to the questionnaire

Gender	Frequency	Percentage	Percentage of responses
Male	201	62.0	65
Female	108	33.3	35
Sub-total	309	95.4	100
Non responses	15	4.6	
Total	324	100	

Table 4 shows the gender breakdown of those who participated in the questionnaire. Interestingly in both runs of the questionnaire (pilot and final) the number of males who completed the survey far outweighed the number completed by females. The most recent available statistics (from the 2010/11 academic year) for the UK shows that 44.2% of academics are female (HESA, 2012). The large difference between this and the percentage of responses provided by female staff in the final questionnaire of 35% means that statistically significant differences between genders could lead to a skewing of the overall picture. This issue is explored in more detail in section 4.2.

Unfortunately, the lack of a 100% response rate does lead to certain issues which have to be taken into account during the analyses of the

available data. It is unlikely that the non-respondents were a random group within the sample (Sheikh and Mattingly, 1981). There is potential for non-response bias in the data as it possible that the academic staff who responded felt differently about the items of the questionnaire when compared with those who did not respond. The respondents were self-selected as they had a free choice about whether or not they completed the questionnaire. Often it is suggested that those with particularly extreme views on the topic of the questionnaire choose to respond. It is not known for certain in this study how those who chose not to respond would have completed the questionnaire. As some of the characteristics of the non-respondents were not observed it is impossible to directly test for non-response bias (Hudson et al, 2004). This is an issue when considering the internal validity of the sample at hand. It is not certain that the analysis of the available data is applicable to the whole of the sample. However, despite this, the large number of respondents does give cause for optimism that at the very least the analysis will provide a useful insight into the views of academic staff across this range of universities.

This chapter has described the development of the approach to this study and the ways in which the effectiveness of the questionnaire has been maximised, both in a practical and theoretical sense. The questionnaire was developed with current practice relating to survey design in mind and this has helped increase the number of respondents and the usability of the questionnaire data. The pilot showed the questionnaire approach to be viable and likely to generate a sufficient number of responses to allow a meaningful analysis of the resulting data and some lessons were learned that improved the final administration of the questionnaire. The analysis of the demographic variables towards the end of this chapter have shown that in a broad sense the academic staff who responded to the survey can be seen as representative of the rest of the sample although caution will be required when claiming the results to be generalisable across the whole of the higher education sector. The next chapter begins to explore the data arising from the questionnaire in more

depth, building a picture of the overall views of academic staff towards the NSS.

[1] The number generator chosen for this purpose is available at http://www.random.org/ [accessed 3 January 2011].
[2] Interestingly this league table assigns points for the NSS scores of each respective institution and the weighting given to "Student Satisfaction" is the joint highest, equal with the A/AS Level UCAS points required to enter the university.
[3] When the target sample was being collated only one job title was recorded per person. Whether or not they were Head of Department was not recorded, although some survey respondents chose to use Head of Department as their job title when answering that question of the survey.
[4] This percentage includes those people whose job titles were not included as part of their departmental web page.

4. Overall results and analysis

As outlined in the previous chapter the questionnaire generated data which could be used to respond to the research questions of this study. This chapter will provide an overview of the results of the questionnaire contributing to the understanding of the ways academic staff perceive the NSS as well as the way the survey is used within their working lives. This in turn will reveal the extent to which the NSS is used for the purposes of quality enhancement.

4.1. Top level results from the questionnaire

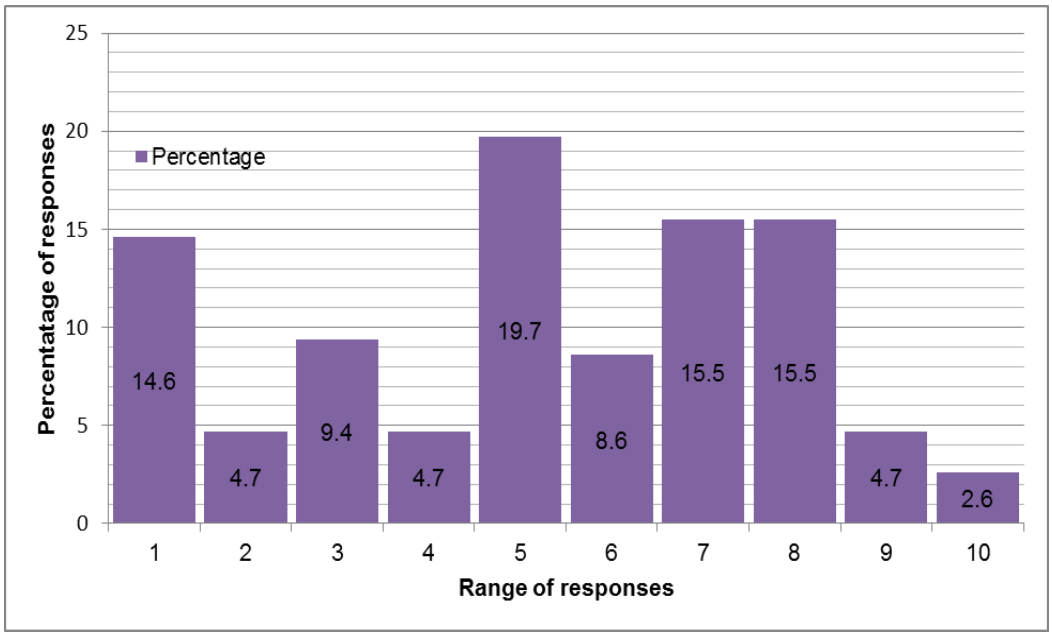
The first statistical test applied to the dataset was the determination of the Cronbach's alpha co-efficient, which for the main 17 items of the questionnaire was 0.850, showing a highly reliable scale. This was a higher result than had been achieved during the pilot questionnaire and this is likely due to the number of respondents. There were good levels of consistency between the variables; in other words the alpha would not have changed much if any one of the items had been deleted from the scale. This provided some confidence that the items from this part of the questionnaire were measuring a similar underlying concept. It is also useful to note that no two variables correlated at a magnitude greater than 0.82, which suggests that collinearity is not an issue in the case of this questionnaire. This is particularly important given the assumptions required to conduct an ordinal regression analysis.

<u>Perceived knowledge about the NSS</u>

Staff were asked to rate their own knowledge of issues around the NSS out of 10, with 10 being the highest rating. There were over 90 non-responses to this item which may have been down to the positioning of the question at the very top of the questionnaire, in a location where respondents were less likely to notice it. However there were a good

number of responses (n=233) and a summary of these is in the chart below

Figure 5: Distribution of responses from academic staff when asked to rate their knowledge of the NSS



The mean response was 5.21 and the standard deviation was 2.581 showing academics to have a wide range of perceived levels of expertise in matters relating to the NSS. It is interesting to note the percentage of people who marked themselves 1/10 compared with the much smaller percentage who gave a rating of 10/10.

Bearing in mind the research questions it was important to include the responses from each survey participant in the wider analysis, even if they felt their knowledge was less extensive. In order to confirm the importance or otherwise of perceived knowledge on the overall perspectives towards the Survey, the respondents to this question were coded into two groups: those who rated their knowledge as five or below and those who rated it above five. The groups were roughly even in size, with 53.2% of those who answered the question in the 1-5 group. Table 6 shows the breakdown of these responses.

Table 6: Distribution of responses for lower and higher perceived levels of knowledge about the NSS

Question	1-5 Disagree %	1-5 Neither agree nor disagree %	1-5 Agree %	6-10 Disagree %	6-10 Neither agree nor disagree %	6-10 Agree %	Sig
Q1	1.6	4.0	94.4	0.9	1.8	97.2	0.549
Q2	1.6	4.1	94.3	0.9	3.7	95.4	0.880
Q3	44.8	46.0	9.2	55.6	25.9	18.5	0.008
Q4	1.1	34.8	64.1	6.7	20.0	73.3	0.015
Q5	17.3	39.8	42.9	19.3	42.2	38.5	0.813
Q6	17.0	35.1	47.9	22.0	35.8	42.2	0.604
Q7	48.5	32.0	19.6	48.6	28.0	23.4	0.741
Q8	73.5	21.6	4.9	59.6	29.8	10.6	0.082
Q9	35.8	37.9	26.3	31.8	33.6	34.6	0.446
Q10	86.1	11.9	2.0	89.4	10.6	0.0	0.333
Q11	23.7	48.4	28.0	20.8	32.1	47.2	0.016
Q12	28.6	42.9	28.6	35.8	43.4	20.8	0.378
Q13	21.8	46.2	32.1	28.8	41.3	29.8	0.558
Q14	17.3	23.5	59.2	18.7	16.8	64.5	0.493
Q15	16.3	16.3	67.4	2.1	9.3	88.7	0.001
Q16	61.7	24.3	13.9	63.0	22.2	14.8	0.926
Q17	40.2	37.1	22.7	44.3	24.5	31.1	0.125

There are a number of items revealing statistically significant differences at the $p<0.05$ level (Q3, Q4, Q11 and Q15). However, none of these questions showed a change in the general direction of the responses. In each of these questions there were a larger proportion of those with lower levels of knowledge answering in the middle of the Likert scale. This was the case for the majority of questions but was apparent to a greater extent in those items with statistically significant differences. This could be because more respondents felt they did not have sufficient expertise to respond with agreement or disagreement. Question 15 asked respondents to state whether their institution had shared results with them. The lower level of agreement and higher level of disagreement within the group of staff with lower levels of perceived knowledge is understandable as they are less likely to have seen results, thus contributing to their lower perceived knowledge. It is also interesting to note that the correlation between individual academics' perceived

knowledge and the overall view as to whether or not the NSS is a useful tool for improving teaching in higher education was very small (0.005) and not statistically significant. This suggests that the level of knowledge held by the academic has little influence on their perception of the NSS for enhancement. It is not the case therefore that those who "know" about the NSS rate it more highly than those who do not or vice versa, suggesting that there are a number of other factors that contribute towards the perception of the NSS as a tool for enhancement.

General perceptions towards the NSS

Respondents had the opportunity to rate their level of agreement with a series of statements relating to issues about the NSS. A score of five, was given to strong agreement and a score of one to strong disagreement. Two tables are below showing a breakdown of the results of these items (specific wording for each of the items is available in Appendix 2).

Figure 7: Summary of responses to Likert scale items – Means and Standard Deviation

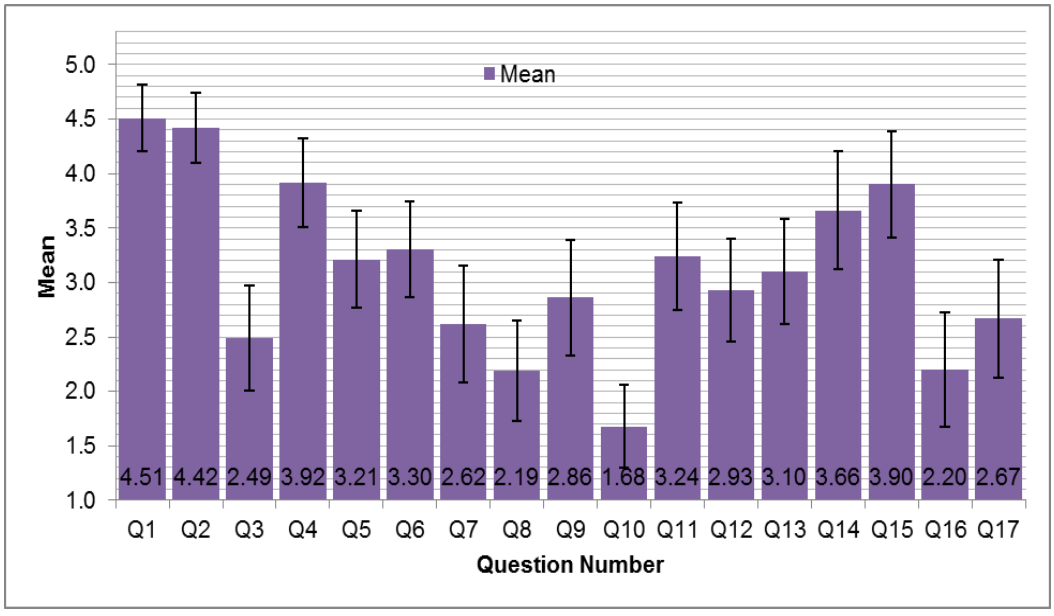


Figure 7 reveals some interesting points about the overall views of the respondents. Firstly there are a number of questions where there

appeared to be relative consensus amongst academic staff as shown by high or low means and relatively small standard deviations. Questions where this is the case include questions 1 and 2 implying that in general the academic community feel that students should have the opportunity to offer opinions on the quality of their course and that these views are important. There seems to be consensus in a negative sense around question 10 which reveals that on the whole academic staff prefer other methods of gathering student feedback over the NSS. This is a point which is illuminated further in the chapter exploring how the NSS is used in practice. Interestingly the question with the widest dispersion of responses was Q17 about the overall view towards the NSS. This added to the interest in conducting an ordinal regression analysis in order to determine which of the other variables contributed to this item.

Figure 8: The levels of agreement with the statements in the Likert scale items. The median response can be seen by looking across the 50% line

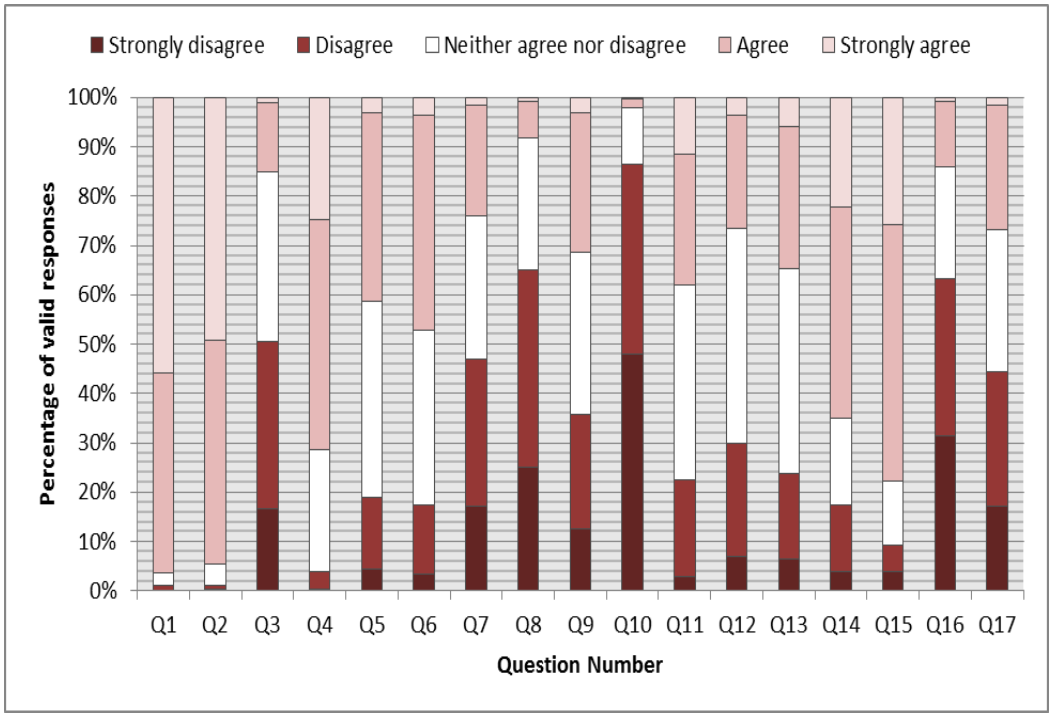


Figure 8 lends even more support to the idea that the scale in use here is inherently reliable. The questions that were intended as conceptual opposites do have different levels of agreement than the other questions. Some questions, for example Q12 and Q13 are far more neutral in

wording and relate to a slightly different notion of departmental and institutional ability to utilise the NSS. It is clear therefore that respondents have been reading the questions and discriminating between the response options. The overall picture being revealed is one which suggests that in general academic staff have relatively negative perceptions of the ability of the NSS to be used as a tool for enhancement. There are low levels of agreement with Q17 with only 15.2% of respondents agreeing with the idea that the "NSS is a suitable measure of teaching quality". The results from these two statements alone are quite damning for supporters of the NSS and this particular finding does directly oppose the viewpoint of some authors who feel the NSS is accepted across the higher education sector, including the view of the Centre for Higher Education Studies (2010). What is potentially of even greater significance is the level of agreement with the idea that there are other more suitable tools for measuring student views (Q4) and the fact that the NSS seems to be more of a concern to senior management than individual teachers (Q14). This latter point is explored in more detail in the chapter on how the NSS is used, but the implication is that the NSS is used to make changes at a level which is removed from the interaction between student and teacher. Relatively few academic staff felt the NSS had made a positive contribution to the development of their teaching, as revealed by the responses to Q8.

The correlations between the main 17 items of the questionnaire proved useful as a way of establishing basic relationships between the variables. There are a number of item pairs which one would expect to correlate strongly (a matrix of the correlations between the items is available in Appendix 5). A few of these correlations require further commentary due to their importance with respect to the research questions. For example it is interesting to note that those who do not believe that the NSS is a suitable measure of teaching quality (Q3) are those who generally feel their own teaching has not improved as a result of using the NSS data (Q8). It is not known however which perception causes which and there may be another unobserved variable proving to be influential. Another

interesting correlation is the inverse coefficient of -0.518 between Q3 and Q11. This shows that those who see the NSS as being a distraction from other methods of improving teaching and learning are generally those who do not think the NSS measures teaching quality in a suitable way. There is another inverse relationship between the idea of high scores showing examples of good practice (Q5) and the distraction caused by the NSS (Q11) with a correlation coefficient of -0.455. This makes logical sense because if high scores are given credence within an institutional context this will lead to attention being paid to these scores. This will be seen as a distraction by those who do not believe these data to be useful. Question 11 also negatively correlates strongly to Q17, and it is understandable that those who do not see the NSS as being useful are more likely to see it as a distraction. What is less clear is whether they simply do not rate the instrument, or whether it is the distraction it causes that makes it less useful for enhancement purposes.

There is a very strong correlation between the items asking whether or not the respondent's department and institution can use the NSS data more effectively (Q12 and Q13), with a correlation of 0.820. This could be due to the interrelationship between the institutional and departmental mechanisms in place to manage the responses to the NSS data. This is explored in more detail in the chapter looking at the qualitative data gathered during the questionnaire (see 4.3). Many academic staff see the level of appropriate response as being the department or institution, rather than the individual staff member. Individuals respond by implementing the changes pressed upon them by others. It is other, internal surveys that seem to promote a more individualised response from academic staff in the view of the respondents to this survey.

Table 9: Ordinal regression analysis of the seventeen core questionnaire items using Question 17 as the dependent variable

Threshold/Location	Estimate	Std. Error	Wald	Sig.
[Q17 = 1]	5.686	1.105	26.504	0.000
[Q17 = 2]	8.766	1.197	53.635	0.000
[Q17 = 3]	11.624	1.326	76.900	0.000
[Q17 = 4]	16.318	1.562	109.161	0.000
Q3	0.608	0.195	9.716	0.002
Q5	1.031	0.223	21.328	0.000
Q16	0.581	0.153	14.467	0.000
Q9	0.731	0.181	16.261	0.000
Q8	0.420	0.226	3.452	0.063
Q7	0.545	0.205	7.024	0.008
Q11	-0.428	0.169	6.419	0.011

The number and range of statistically significant correlations between the first 16 items of the questionnaire and the overall question about the usefulness of the NSS suggested that this required further investigation by means of a regression analysis. Regression methods are a good way of studying the relationship between an output variable and the input variables as they take into account the interrelationships between the variables. An ordinal regression model is required as the dependent variable, in this case question 17 of the questionnaire, uses a Likert scale and is therefore ordinal (Chen and Hughes, 2004). A model was fitted in a stepwise fashion using the logit link as this was the function that produced the necessary result in the test of parallel lines, which is an important requirement of the model (SPSS, 2002). When fitting this model the variables were added as covariates because the Likert items were deemed to be more like continuous, rather than categorical variables.

This regression model was significant at the $p<0.001$ level and the Nagelkerke pseudo R^2 value is 0.720. This value shows the difference between the model's estimation and the actual results in the dataset (Veall and Zimmermann, 1996). Table 9, showing the contribution of each of the items to the model is above. The figures in the estimate column show the contribution the item makes to the probable outcome of Q17, so for example, an increase of one in the response to Q5 would increase the

probable response for Q17 by a factor of 1.031. The relationships shown are all significant at the $p<0.05$ level with the exception of Q8 which is significant at the $p<0.1$ level. The four rows relating to Q17 at the top of Table 9 shows the model to be statistically significant at each of the cut points determined by the model (which in this case are the range of responses offered to respondents). This is another indication of the suitability of this model for predicting across the range of responses for Q17.

The stepwise method employed in building this regression model meant some of the variables were excluded as they did not improve the effectiveness of the model in predicting the dependent variable. This therefore highlights the importance of certain variables and the way these contribute to the overall perceptions of the NSS. The items included in the model focus on the suitability and usefulness of the NSS as both an enhancement tool (for example Q8 “I think that my own teaching has improved as a result of making changes informed by NSS data”) and as a general performance indicator (for example Q3 “The NSS is a suitable measure of teaching quality”).The items not contributing to the final regression model focus on the idea of gathering data from students *per se* (for example Q2 “Students should have an opportunity to rate the quality of their course”); the comparison between the NSS and other tools, such as Q10 asking respondents to state whether the NSS was their preferred feedback option and the mechanisms by which NSS data are used within institutions and departments (an example being Q12, “My department/faculty could use the NSS data more effectively than it currently does”). However as we see elsewhere in this study, the issues raised by these items remain important, particularly when considering the relationship between the NSS and other survey tools used for improving teaching and the ways in which Universities manage their processes to respond to the NSS. It may be that when considering their response to Q17, respondents were emphasising some considerations above others before stating their level of agreement.

This overview of the quantitative data gathered through the questionnaire has shown the research instrument to be reliable. With this in mind it can be reasonably suggested that there are issues with the NSS as perceived by this particular group of academic staff. It is not clear from the quantitative data alone what the causes of those issues are. The data reveals a substantial amount of negativity towards the NSS with only small minorities seeing the survey as a useful tool which provides them with meaningful and usable data. What is perhaps more concerning is the implicit notion that the NSS might actually be preventing meaningful enhancement work and the very small percentage of staff who see the survey as having contributed to a positive change in their teaching.

4.2. Analysis by gender

Although a specific analysis by gender was not one of the research questions of this study, it was shown earlier in the thesis that there was a discrepancy between the percentage of respondents of each gender in the sample and the percentages of staff of each gender in the sector as a whole. This could be a potential source of bias in the sample as the large percentage of male respondents is not representative of the population. The questionnaire was completed by 201 male academic staff compared with 108 female academic staff.

With this in mind an analysis of the responses by gender would allow any effects of this bias to be identified as well as showing the differences (or otherwise) between the genders when considering the National Student Survey. The table below provides a breakdown of the responses to the core items of the questionnaire disaggregated by gender.

Table 10: Distribution of responses for each gender

Question	Male Disagree %	Male Neither agree nor disagree %	Male Agree %	Female Disagree %	Female Neither agree nor disagree %	Female Agree %	Sig
Q1	1.0	3.5	95.5	1.9	0.9	97.2	0.334
Q2	1.0	4.0	95.0	1.9	4.7	93.5	0.776
Q3	50.9	30.5	18.6	47.4	41.2	11.3	0.123
Q4	5.8	24.3	69.9	0.0	25.0	75.0	0.056
Q5	18.2	38.6	43.2	19.4	40.8	39.8	0.858
Q6	16.1	31.6	52.3	19.8	38.6	41.6	0.230
Q7	45.1	28.6	26.4	47.9	30.2	21.9	0.711
Q8	65.2	26.0	8.8	64.3	27.6	8.2	0.950
Q9	32.0	34.8	33.1	40.6	28.1	31.3	0.325
Q10	86.5	10.8	2.7	87.1	11.8	1.1	0.664
Q11	27.6	33.3	39.1	15.2	50.0	34.8	0.014
Q12	27.4	46.3	26.2	32.3	39.6	28.1	0.549
Q13	21.4	44.7	34.0	26.7	37.8	35.6	0.503
Q14	15.1	16.8	68.2	19.4	19.4	61.2	0.492
Q15	10.3	11.6	78.1	8.1	16.3	75.6	0.541
Q16	63.7	21.2	15.0	62.4	25.7	11.9	0.583
Q17	43.8	28.1	28.1	45.9	28.6	25.5	0.895

This analysis reveals only minor differences between the genders. Not only are the differences in the percentages against each response quite small but the levels of significance are at the $p>0.05$ level. These findings suggest that the perceptions towards the NSS are only marginally affected by gender, meaning that the potential bias caused by the larger proportion of men completing the survey does not seem to have created a slant in the overall results to the questionnaire. This is a helpful finding when intending to make overall conclusions about the perceptions of these academic staff towards the National Student Survey.

Understanding the lack of difference between the genders is an interesting finding in itself. Perhaps the National Student Survey is not being viewed through a gender related lens and the perceptions staff have are affected by other factors when making their judgements about the NSS.

4.3. How the NSS is used

Respondents to the questionnaire were asked questions seeking to find out how both they and their department used the results of the NSS. They were asked specifically how their department or faculty used the NSS and why, as well as how they, as an individual used the survey. Over 80% of questionnaire respondents answered these questions in some form or another. There was also a multiple select question which asked each person completing the survey to suggest where the motivation for using the NSS comes from in their working environment.

The gathering of the qualitative data was primarily aimed to provide data for the research question on how academic staff use the NSS for enhancement. This qualitative data needed to be analysed in a different way, as part of the mixed methods approach of this study. For the purposes of analysing the data, NVivo 7 was used and a grounded theory approach was employed (Glaser and Strauss, 1967). Numerous initial codes were developed which were then collated to develop broader themes. Further analysis of the collected data is below.

Awareness of the NSS

There were some issues with the answers provided by some of the respondents. These usually involved the respondents declaring their response invalid for one reason or another. These reasons typically included not actually being teachers of undergraduate students, usually because of a largely postgraduate teaching portfolio. There were however another group of people who declared that they were not in a position to answer the questions, using a reason that is far more relevant to the research questions of this study. A number of academic staff felt that their awareness of issues relating to the NSS was slight, in other words they were not sure about how the NSS was used within their department or faculty, even if it would be relevant for them to know. This was sometimes related to intra-departmental communication or the ways the results were

presented to staff. It could be argued that awareness of the NSS is the first step to using it for enhancement purposes. It seems that effective communication of the results and their implications is an important prerequisite for the use of the data.

> I would like to know more about it and have more information. The results which are fed back to us tend to be of a general nature, so they are not easy to use as individuals. (Uni 2, History, Other)

> I have never really been given any data from the survey to use. (Uni 7, History, Lecturer)

Perhaps related to this issue of communication was the claim from a number of academics that the NSS was not used at all, either by their departments, or by the individual member of staff. A number of people answered the question “How do you, as an individual, use the results of the NSS? Why?” with “I don’t” or similar. Within the departmental context it was not always the case that people were aware of how the survey results were used.

> I am not aware of any use of the NSS being made in either my department or faculty. (Uni 12, History, Senior Lecturer)

> I have seen no evidence that the NSS is used at all. (Uni 3, History, Lecturer)

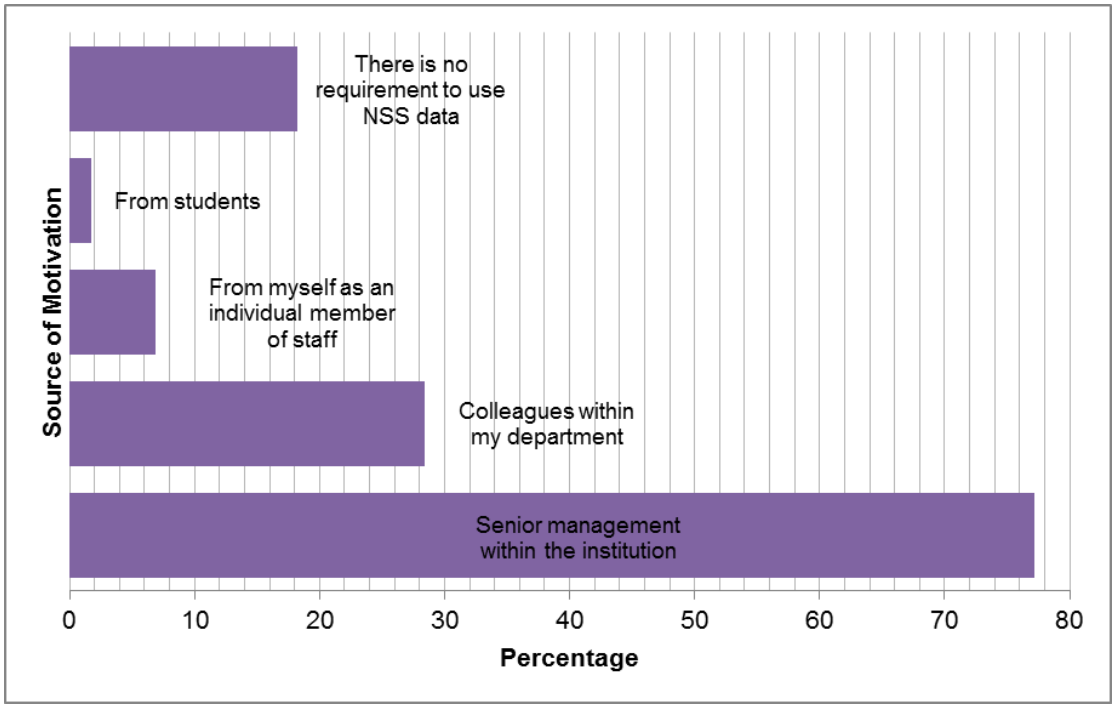
> Never heard of it before now! (Uni 3, Physics, Reader)

Relationship with senior management

The importance of relationships within institutions has emerged as a key theme from this study. The quantitative results arising from the questionnaire distributed to academic staff indicated that, in the views of the teachers themselves, issues relating to the NSS were of more concern to senior management than to teachers, with 64.9% agreeing

with that statement. It is also clear that in a number of contexts, the impulse to respond to the results of the NSS comes from senior management, as indicated by the table below. This table shows the results from a multiple select question asking academics to suggest where the requirement for them to act in response to NSS results is usually from.

Figure 11: Responses to the multiple select question about requests to act upon NSS results



As can be seen from the figure, the majority of people responding to this question cited senior management within their institution as being those who request further action based on NSS results. The reaction of senior management to the NSS results and the way in which this manifests itself in institutional policy and process could have an impact on the way the NSS is used for enhancement. Unfortunately however, academic staff occasionally cited less than positive relationships with parts of the university outside their own department or faculty. The language being used by the survey respondents was often based on a “them and us” dichotomy that seems to breed a type of resentment. Academic staff would talk of the need to be responding to the NSS in some form or

another to please senior management. Others suggested that the way senior management respond to the survey contributes to a general feeling of low morale.

> Mostly as something to worry about and to use as a way of transferring blame from idiots in serionr [sic] management to teaching staff. (Uni 12, Other subject, Lecturer)

> As yet another stick for university managers to beat us over the head with (to go with the QAA, RAE, REF, 'Impact', TLHEP, pressure to bring in research funding, etc., etc.). I wish that there was a National Staff Survey to go alongside the National Student Survey! (Uni 9, History, Lecturer)

> To use as a weapon against us; to get what they want! which mostly is for us to fill in another 6-page form, which in my institution's philosophy seems to be the answer to all ills. (Uni 2, History, Lecturer)

There is a perception amongst the respondents to this questionnaire that the primary aim of the senior managers within their institutions is simply to improve the raw scores in the NSS, rather than to enhance the learning and teaching experience of students. There was no indication of a more nuanced strategy or partnership between academic staff and their senior counterparts which was co-designed in order to develop meaningful enhancement activities; again this is a function of the “them and us” paradigm expressed above.

> To improve feedback as that is what managment [sic] think needs to be addressed. (Uni 10, History, Lecturer)

> Attempting to avoid possible negative feedback due to institutional / managment [sic] use as 'improvement' and efficiency tool. (Uni 2, Education, Senior Lecturer)

It is also assumed that the motivation for improving the raw scores comes from the need of the institution to perform well in institutional league tables. League tables were mentioned a number of times in the qualitative comments. One of the items in the closed question section of the questionnaire showed that only 13.3% of respondents agreed that league tables were a positive development in higher education. This item also featured as part of the ordinal regression model fitted stepwise showing the influence this item has on the overall views of academic staff towards the NSS. This is a potential explanation as to why academic staff resent the emphasis being placed on improving raw scores if this is the underlying motivation for doing so, rather than a seemingly more noble desire to improve what students are being provided with.

> To see what we need to improve upon to raise the position in the league table. However evidence of use of league tables elsewhere e.g. league tables in schools shows that they promote abnormal behaviour and generate unintended consequences. (Uni 9, Education, Senior Lecturer)

> It [the NSS] is used as a tool to get the university higher in national league tables and that's it. (Uni 2, Education, Lecturer)

> How: To identify perceived issues in student provision. Why: Partly because of the wish to improve student teaching, but like all other Universities it is partly because they are afraid of poor scores in league tables. (Uni 10, Physics, Professor)

In order to establish some of the motivation for feeling positive or negative towards league tables a variable was added to the dataset with the position of the respondent's institution in the 2010 *Sunday Times* league table. The individual responses were coded into three groups: "high", "middle" and "low" depending on their institution's relative league table position. Each of the groups contained the responses from four of the twelve institutions, explaining the slightly different number of respondents in each of the categories. A cross tabulation of these groups

against the item of the questionnaire relating to league tables is contained in Table 12.

Table 12: Crosstab between league table position and response to the item "League tables are a positive development in Higher Education"

Response	Data type	High	Middle	Low	Total
Disagree	Count	48	52	71	171
	% within section of league table position	52.2	62.7	72.4	62.6
Neither agree nor disagree	Count	21	21	20	62
	% within section of league table position	22.8	25.3	20.4	22.7
Agree	Count	23	10	7	40
	% within section of league table position	25.0	12.0	7.1	14.7
Totals	Count	92	83	98	273
	% within section of league table position	100	100	100	100

The cross tabulation shows large and statistically significant ($p<0.05$) differences in perceptions of league tables between those relatively high, middle and low in those rankings. The difference in the extent of disagreement is over 20% between high and low ranking respondents with a 17.9% difference in the level of agreement. Those who are higher in the institutional league tables seem to have a more positive view towards the existence of those tables.

Combined with this is the perceived need of the institution to maintain or improve its reputation to aid student recruitment.

> Advertising our position of strength to the outside world, since we recognise that prospective applicants pay attention to NSS league table position. (Uni 5, Physics, Senior Lecturer)

> As my institution does well in the NSS, it is usually used as a marketing tool (and it it [sic] is true that it is important for us in terms of recruitment). (Uni 6, Other subject, Lecturer)

The relationship between different parts of the institution seems to be an important factor in determining the attitude of academic staff towards the NSS. It could be concluded that as it currently stands, these relationships do not appear particularly positive. This seems to create a problem for those seeking to use the NSS for an enhancement agenda, as the requirements of senior management staff are perceived as being somewhat different to this.

Standard departmental procedures

There was a great deal of evidence suggesting that the NSS results were built into departmental process and procedures. This allows discussions to take place about the nature of the results and the actions needed to respond to the issues raised. It appears to be common practice for departments to compare themselves with other departments within their institution and similar departments in rival universities. If the results are positive this is appreciated.

> To make comparisons between institutions and different departments in the university. I suppose they do this because they feel that this is a worthy thing to do and because they believe that this is what others will be doing as well. (Uni 6, Education, Director)

> To see where our ratings compare with those of otehr [sic] departments in the institution, and with others in our discipline. (Uni 7, History, Professor)

There was no indication of a general awareness of the importance of ensuring that these comparisons are meaningful in a statistical sense. However the practice of comparing like-for-like departments and departments within an institution does broadly chime with the recommendations of Marsh and Cheng (2008) highlighted above. What is less clear is whether departmental staff appreciate the nuances of using the NSS for comparison in this way.

The NSS data is clearly an input into discussions within the department about teaching and learning. A number of respondents to the questionnaire mentioned forums where the NSS is used as the basis for discussion. Meetings mentioned included learning and teaching committee meetings; faculty meetings and discussions within course teams. It appears as if these discussions feed into more concrete action plans that would be implemented across the department.

> Generally, the NSS results are used for another bout of "could do better"-type analysis and navel gazing on the part of the head of school and some faculty. Emails come round from the head of school noting what our score is and how it has changed, and these results are also discussed at school meetings and meetings of teaching committees (Uni 9, History, Lecturer)

> Results are reviewed in Faculty committee and then action is requested from Departments and Schools. (Uni 5, History, Senior Lecturer)

> The results of the NSS are discussed at the Departmental level. If particular issues are flagged up, measures are taken to resolve them. Both the numerical scores and the individual comments are useful in this respect. (Uni 6, Physics, Senior Lecturer)

The NSS seems to be included as an information source as part of the annual cycles of departments, for example during course reviews and annual teaching reviews. The NSS appears to have been firmly built into the quality assurance processes within the institution, possibly because of

the influence of senior management. The quantitative nature of the NSS scores encourages an emphasis on score improvement, rather than the generally softer notion of enhancement. It is not clear whether the inclusion of the NSS as part of these processes is of benefit to teaching and learning. However the fact that it is increasingly being used in this way is apparent from the responses provided by academic staff, as demonstrated by the quotations below.

> To review courses and enhance the profile of 'quality assurance and enhancement' agenda and to increase administration of courses. (Uni 2, Education, Senior Lecturer)

> I use it [the NSS] to review, with colleagues on the teaching programme concerned, areas in the survey where it is clear the results could be stronger. (Uni 4, Education, Professor)

> Mandatory to consider it at the annual course review. (Uni 4, Physics, Professor)

Within these procedures departments generally seem to work with a deficit model i.e. they are looking to find issues rather than discover what they are particularly good at. This deficit model seemed to create a requirement to take some form of action, even if that action was largely unrelated to the issue raised by the survey in the first place. It would be a typical requirement of an action plan to make some changes, using the survey as justification. Feedback was mentioned a number of times as an issue that a department has attempted to address although curiously no other question or scale was mentioned by name (again this may indicate the relatively unsophisticated data analysis employed within institutions). As a result of this process, departmental staff appear willing to implement changes to policy or practice in an attempt to improve scores.

> We may change practice, e.g. introducing additional contact hours in response to management's perceptions and analysis of students'

> complaints (even when we think their comments unjustified and/or ignorant). (Uni 9, History, Professor)

> Identify weaknesses in teaching practice and take mitigating action. NSS provides a good overview from students having the "full" experience and thus may be in a better position to rate the quality of teaching than say first year students with limited exposure to HE. (Uni 5, Physics, Senior Lecturer)

The implication of this is that staff do what is required to comply with institutional process as the expense of a deeper engagement with the survey results. The emphasis is on correcting perceived faults; unfortunately however, these efforts are sometimes misdirected. These issues are a potential symptom of the mismatch between the desires of academic staff and their managers at a senior level.

Use of the data by individual staff

As we have seen from the ordinal regression analysis, the views of individual academics on the ways they have personally benefitted from the NSS seems to be influential on their overall views of the survey. Members of staff offered a wide range of perspectives on the ways they used the NSS data as part of their work on improving their own teaching. There were some respondents to the questionnaire who did not know what the NSS was, and at the other end of the scale there were those who used the NSS in a sophisticated way to inform changes to what they provide. It is not possible in a thesis of this scope for all of the strands to be mentioned but some recurring themes can be investigated.

The NSS is used by many academic staff as a tool for reflection or self-evaluation. Although the NSS does not provide information about individual modules this does not seem to exclude using the survey for this purpose, particularly when the survey results are used in conjunction with more tailored interventions, for example discussing a module with

students or using the results of end of module evaluations. Although the notion of reflection seems vague in a lot of the comments made, it can be said that the NSS is acting as a starting point for this reflection, possibly due to the high-profile nature of the tool.

> I use it to reflect upon whether or not areas of my own teaching fit the departmental profile, and correlate the departmental picture with internal module evaluations (Uni 4, History, Professor)

> I feel that as a teacher and an academic we have to practice [sic] what we preech [sic]. If we're telling students to reflect on experiences and feedback, we have to do the same. It enhances the whole teaching and learning experience - for both tutor and student. (Uni 5, Education, Teaching Fellow)

> I try to understand it as best I can, given the politics and agenda of its intention and design, and use it as part of the way I understand and undertake my work within the University (Uni 8, Education, Reader)

This perhaps indicates the general willingness of academic staff to improve their teaching. The NSS occasionally plays a part in reflection, although this is by no means widespread, judging from the percentage of respondents who agreed with Q7's statement "The NSS provides useful information to help me improve my teaching" (23.7%).

In the questionnaire there were two Likert scale items exploring the idea of other student feedback tools being used instead of the NSS (Q4 and Q10). Over 70% of respondents agreed that there were other more appropriate tools than the NSS and only 2% agreed that the NSS was their preferred method of gathering feedback from students (see Figure 8). These responses were supported by the qualitative comments, with over seventy references to the use of other methods of gathering feedback. There were two distinct perspectives amongst academic staff about the desired interplay between NSS scores and these internal mechanisms. One view was that the module evaluations could be used in

conjunction with the NSS to provide a larger set of data which would help indicate issues. In these examples, the NSS appears to play the role of junior partner in that although it is useful, it is less relevant than the module evaluations to that individual member of staff.

> To monitor student opinion and internally to ensure the depts in the faculty are contributing - however there are more detailed evaluations used within programmes that are more useful in developing the programmes and addressing student learning needs (Uni 5, Education, Professor)

> But NSS results are not sufficiently fine-grained to secure accurate understanding of some of the ratings, so cannot be the sole basis for student feed-back. (Uni 4, History, Lecturer)

The other perspective strongly favours the module evaluations to the extent that they are used in preference to any engagement with the NSS. This perspective was expressed more forcefully than the view that the NSS could play a role. It is interesting to note that the perceived opportunity cost of engaging with the NSS as a tool for enhancement seems to be the use of the other module surveys. This could be a symptom of the multi-faceted roles which UK academics hold within their universities and the pressures they face in other aspects of their employment. It seems to be the case for many staff that engaging with the NSS is a luxury, whilst module evaluations are a necessity.

> I can say that feedback on individual modules or individual teachers carried out within the institution is far more useful and is much more likely to influence teaching. (Uni 9, History, Lecturer)

> Much more significant to the improvement of teaching (and the incorporation of student views therein) are our internal module reviews, where we largely assess the qualitative comments of students on a particular course or module. (Uni 8, History, Senior Lecturer)

As explained in the literature review chapter, the original intention was for the NSS to provide information to the public rather than perform the function of an enhancement tool. A survey of academic staff led to the conclusion that a national survey of students would add little to internal feedback mechanisms as the results would be too general (CHERI et al, 2003). The qualitative evidence from this study closely supports this viewpoint. The perceptions of academic staff would need to change in a significant way to ensure the more widespread use of the NSS for enhancement purposes.

However, the above point does not mean that academic staff are unwilling to change their practices because of outcomes from the NSS. It is merely suggesting that academic staff do not generally see the relevance of the NSS to their own teaching. A number of academic staff said they responded to the NSS by implementing any departmental-wide changes deemed appropriate following discussions amongst colleagues.

This chapter has explored the data collected via the questionnaire at the aggregate level to determine the views of staff in general towards the NSS and establish the ways in which the survey is used for enhancement. There are some key themes emerging from the analysis of both the quantitative and qualitative data. It is clear that there is some apathy towards the NSS, particularly when contextualised within the operations of institutions, where league tables are regarded as an issue and senior managers are deemed to have unhelpful agendas. The qualitative information suggests that the NSS data is used, but rarely in isolation from other forms of student data and the preference is to primarily use internal feedback because of its relevance to the context. The next chapter will seek to establish the nature of the differences in views between staff from different disciplinary backgrounds as well as differences across institutional types.

5. Disciplinary and Institutional Levels

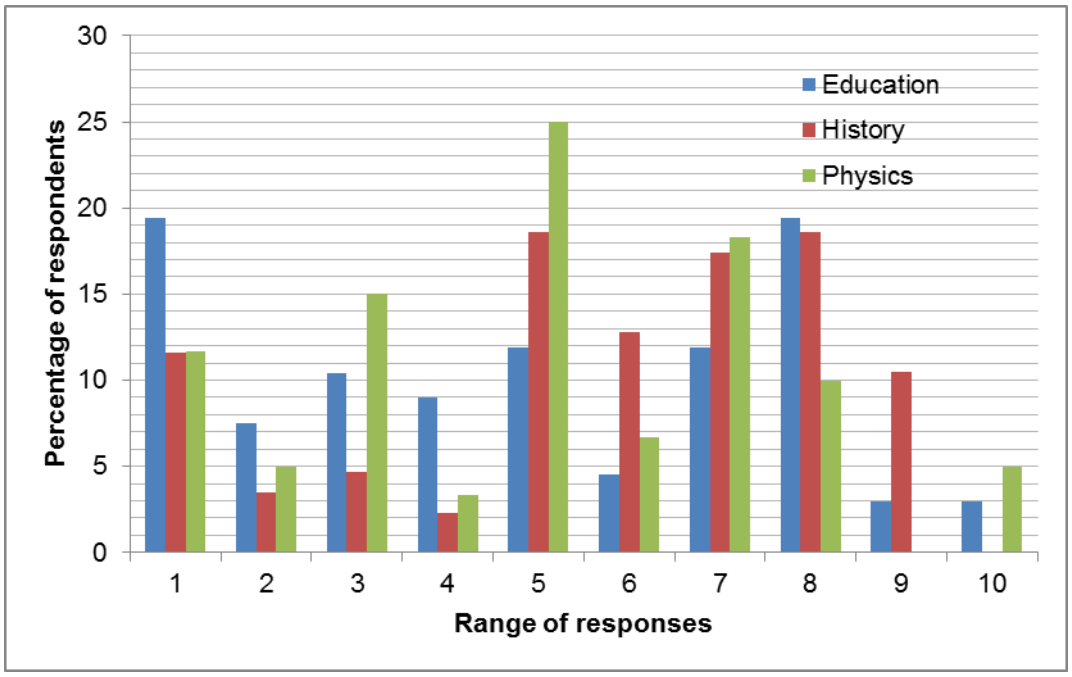
As described in the section on disciplinary differences there is an often articulated view that academic staff identify themselves primarily with their discipline over and above any other conception of identity that they hold. Within each discipline there are characteristics that make the area different, one of the identified differences being the level of affinity certain disciplinary areas have with interventions designed to foster pedagogic improvements (Braxton, 1995). One of the research questions of the present study relates to the existence or not of this phenomenon when the NSS is the teaching and learning intervention in question. This chapter will explore the perceptions of academic staff from three disciplines: Education, History and Physics, which each sit in separate areas of the typology developed by Biglan (1973). As is shown in Table 2 above, academics from each of these disciplines completed the questionnaire in sufficient numbers to make meaningful comparison possible. This chapter will establish the nature of the differences in perceptions between staff of these subject areas; the extent and importance of these differences and discuss the implications of the findings.

In addition this chapter will also explore differences at an institutional level to establish any patterns in the data. As stated above, an academic's primary affinity is with their discipline, however due to the nature of the NSS and its importance in the construction of league tables (and the assumed reputational impact of these) it is interesting to note any significant differences between institutions of certain types. As can be seen in Table 1, academics from twelve universities completed the questionnaire, providing useful data on this topic.

5.1. Disciplinary differences

As with the analysis of the whole dataset, descriptive statistics were produced to establish the magnitude of the differences between the respective groups of academic staff. The first question analysed was the level of NSS related knowledge the staff assessed themselves as having. The results of this disaggregation are summarised in the figure below.

Figure 13: Disaggregation by discipline of responses to the question asking academic staff to rate their knowledge of the NSS

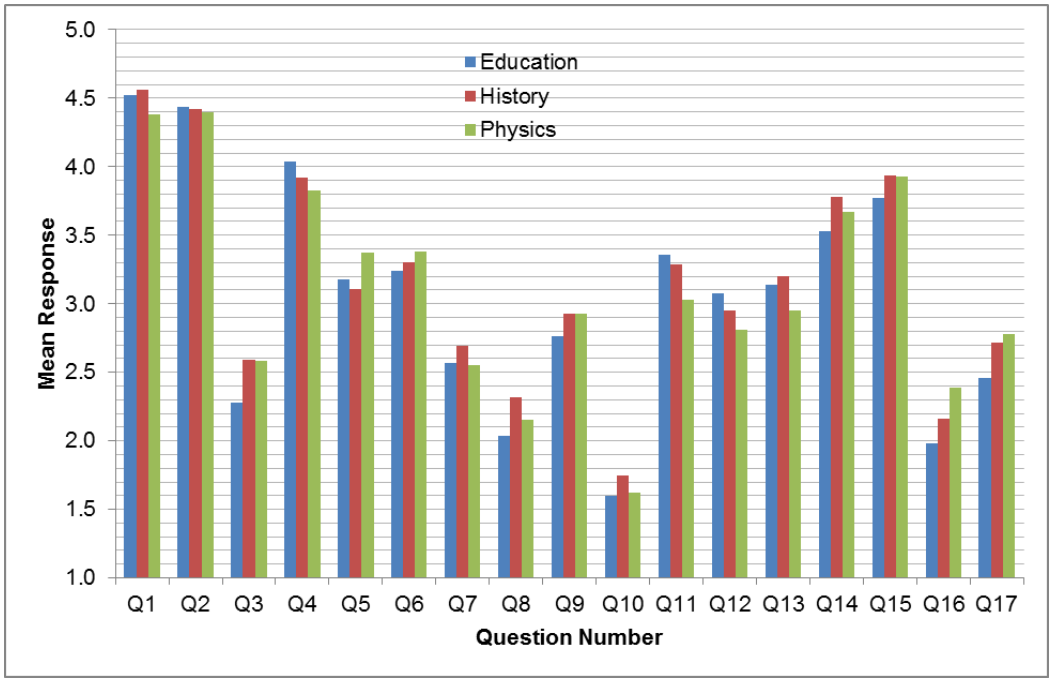


The highest mean response was from historians (5.77) compared with 5.03 for Physics and 4.84 for Education. The largest standard deviation was provided by the Education academics (2.805), compared with 2.458 for History and 2.449 for Physics. As you can see from Figure 13, historians tended to group their responses in the 5-8 bracket with very few responses in the 2-4 range. They achieved the highest mean despite the fact that no History academic rated themselves 10/10. Education academics used the full range of possible responses and had a high percentage answering 1/10. Although physicians used a wide range of responses, the high percentage who answered 5/10 contributed to a

mean of 5.03 and a relatively small standard deviation of 2.449. This figure shows that within each of the disciplinary groups there were academics who rated themselves highly and those who were less confident of their knowledge. There is not a discernible pattern of one group of academic staff seeing themselves as being consistently and substantially more knowledgeable about matters relating to the NSS.

The Likert items within the questionnaire provided an opportunity for academic staff to rate various statements relating to the NSS. The responses to these 17 items are disaggregated by discipline in the figure below.

Figure 14: Means for the Likert scale items, disaggregated by discipline



In general terms there do not appear to be any major differences in the distribution of the responses when they are disaggregated by discipline. The largest difference between two means for a particular question is 0.41 for question 16 between Education and Physics. In a five point scale these differences are not of major practical importance, although they are of interest. The standard deviations for each of the questions also show a great deal of similarity between the disciplines. Although the responses

are not normally distributed these are helpful when determining the level of agreement within a sample of respondents. There are only very slight differences between the subject areas when describing the level of consensus for particular questions.

Crosstabs between the subject areas and the level of agreement with the statements in the question largely ratify the picture above. However there were some exceptions, where items from the questionnaire revealed statistically significant differences between the distributions of responses. The three items where this was the case were Q13 (My institution could use the NSS data more effectively than it currently does); Q15 (My institution shares results with individual departments/faculties) and Q16 (League tables are a positive development in Higher Education). The distribution of responses to these three items resulted in statistically significant differences at the $p<0.05$ level. All three of these items have an institutional level dynamic to them and so it is a slight surprise to see differences in the responses between academics of different disciplines. However, because they are questions that could be regarded as institutional in nature, the differences between the subjects are questionable and not conclusive.

Although the top level disciplinary analysis only revealed minor differences in the distribution of responses when subjects were compared with each other it was considered as useful when answering the research question to compare each subject with the responses for all other subjects. This would be another way of discovering any inherent differences in the viewpoints of academics from different disciplines. In order to achieve this, the dataset was recoded three times with each subject being isolated from the rest of the data. This does make an assumption that the two other subjects were acting as representative of all other disciplines but these data were only being used in an indicative sense.

Table 15: Comparison of levels of agreement between Education and the other subjects combined

Question	Education Disagree %	Education Neither agree nor disagree %	Education Agree %	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Sig
Q1	2.1	1.0	96.9	0.9	3.1	96.1	0.382
Q2	1.1	5.3	93.7	1.3	3.9	94.7	0.855
Q3	57.7	34.6	7.7	47.5	34.3	18.2	0.074
Q4	0.0	24.7	75.3	5.5	24.5	70.0	0.096
Q5	20.0	37.6	42.4	18.4	40.8	40.8	0.878
Q6	17.9	35.7	46.4	17.3	35.1	47.5	0.985
Q7	49.4	27.7	22.9	45.9	29.5	24.6	0.864
Q8	70.7	23.2	6.1	62.7	28.2	9.1	0.409
Q9	39.0	31.7	29.3	34.3	33.3	32.4	0.745
Q10	87.8	12.2	0.0	85.9	11.2	2.9	0.291
Q11	23.1	35.9	41.0	22.1	41.2	36.7	0.705
Q12	26.0	36.4	37.7	31.4	46.4	22.2	0.033
Q13	25.0	34.7	40.3	23.4	44.1	32.4	0.351
Q14	27.1	10.6	62.4	13.3	20.7	66.0	0.006
Q15	19.2	6.8	74.0	5.0	15.6	79.3	0.001
Q16	71.1	18.9	10.0	60.0	24.2	15.8	0.170
Q17	50.0	29.3	20.7	42.2	28.4	29.4	0.292

When Education is compared to the other responses there were three crosstabs showing statistically significant differences at the $p<0.05$ level as can be seen from Table 15. These differences were for questions 12, 14 and 15. Question 12 asked respondents to rate whether or not their department could use the NSS more effectively than it currently does. The Education academics agreed with this more often than their counterparts from other subjects. Fewer Education academics responded in the middle of the scale, suggesting an extra confidence in stating their view on the way the department uses the survey, whether they know much about the survey or not. This is validated by the lack of significant differences for Q12 when History and Physics were compared to the other responses (showing significance at $p<0.718$ and $p<0.148$ respectively). Question 14 asked respondents to rate the extent to which the NSS was more of a concern to senior management when compared with departmental staff. Although levels of agreement were roughly the

same, the extent to which Education academics disagreed with the statement compared to other academics suggest that at least a rump of Education academics are concerned with the NSS as much as the senior managers are perceived to be. With Q15 the main difference in the distribution of responses was seen in the level of disagreement with the statement. Education academics more often felt that their institution did not share the results of the survey, although the percentages are notably low. This could be a point about the level of sophistication of the results as presented by the institution. Staff within Education departments are possibly more likely to understand the nuances of the ways these data are presented and therefore disapprove if this is done in an overly simplistic way.

Some of the differences between the views of Education academics and the rest of the sample are interesting and perhaps hint at a different perspective from this group. However the overwhelming picture is one of similarity with the other subject groupings.

Table 16: Comparison of levels of agreement between History and the other subjects combined

Question	History Disagree %	History Neither agree nor disagree %	History Agree %	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Sig
Q1	0.9	1.7	97.4	1.4	2.9	95.7	0.716
Q2	0.9	3.4	95.7	1.5	4.9	93.7	0.739
Q3	45.3	34.0	20.8	53.5	34.7	11.8	0.113
Q4	6.7	20.0	73.3	2.3	27.3	70.5	0.094
Q5	21.3	45.4	33.3	17.5	36.6	45.9	0.109
Q6	18.7	36.4	44.9	16.8	34.6	48.6	0.817
Q7	43.0	33.6	23.4	49.2	26.2	24.6	0.392
Q8	60.4	27.9	11.7	67.8	26.1	6.1	0.194
Q9	34.6	29.9	35.5	36.3	34.6	29.1	0.496
Q10	83.8	12.4	3.8	88.0	10.9	1.1	0.270
Q11	21.4	36.9	41.7	23.0	41.4	35.6	0.594
Q12	27.4	46.2	26.4	31.5	41.8	26.7	0.718
Q13	19.4	40.8	39.8	26.8	42.0	31.2	0.252
Q14	13.1	19.6	67.3	19.9	16.6	63.5	0.318
Q15	5.2	15.6	79.2	11.5	11.5	76.9	0.182
Q16	63.4	27.7	8.9	63.2	19.7	17.1	0.069
Q17	44.3	30.2	25.5	44.4	27.8	27.8	0.875

When the responses from History academics were compared with the other responses it revealed no statistically significant differences at the $p<0.05$ level. This is a surprising finding in itself but does lend support to the idea that there is in general an overarching perspective on issues relating to the NSS with History academics generally adhering to the views espoused by the academic community as a whole.

Table 17: Comparison of levels of agreement between Physics and the other subject areas combined

Question	Physics Disagree %	Physics Neither agree nor disagree %	Physics Agree %	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Sig
Q1	0.0	6.3	93.8	1.6	1.2	97.1	0.023
Q2	0.0	3.8	96.3	1.6	4.5	93.8	0.487
Q3	49.3	32.8	17.9	50.7	34.9	14.4	0.776
Q4	5.7	27.1	67.1	3.3	23.7	73.0	0.528
Q5	13.7	37.0	49.3	20.6	40.8	38.5	0.208
Q6	13.9	34.7	51.4	18.7	35.5	45.8	0.583
Q7	50.0	25.7	24.3	45.8	30.1	24.1	0.749
Q8	66.7	25.0	8.3	64.4	27.4	8.2	0.923
Q9	33.8	35.2	31.0	36.3	32.1	31.6	0.880
Q10	88.2	10.5	1.3	85.8	11.8	2.4	0.817
Q11	25.0	45.8	29.2	21.5	37.6	41.0	0.204
Q12	32.8	50.0	17.2	29.0	41.5	29.5	0.148
Q13	27.4	51.6	21.0	22.7	38.4	38.9	0.033
Q14	12.5	25.0	62.5	19.0	15.3	65.7	0.118
Q15	5.0	20.0	75.0	10.4	10.9	78.6	0.110
Q16	59.2	19.7	21.1	64.6	23.6	11.8	0.128
Q17	40.5	27.0	32.4	45.8	29.2	25.0	0.459

The comparison between Physics and the other subject areas revealed two statistically significant differences at the $p<0.05$ level, for questions 1 and 13. The result for Q1 can largely be put down to the absence of any responses within the “Disagree” categories from Physics academic staff. This question, which asked respondents to state their agreement with the notion that student views are important, received a high level of agreement from both groups meaning the differences are of little practical significance. The differences shown in Q13 are more interesting, with a high percentage of physicians choosing to respond in the middle of the scale and a lower percentage choosing to agree. This item asked respondents to rate the effectiveness of their institution’s use of the NSS data. This could be due to a difference in the way Physics departments engage with institution-wide work relating to the NSS. Perhaps the more structured nature of Physics departments allows for less flexibility in engaging or learning about the NSS through institutional activity.

The tables above actually show a notable similarity between the subject areas. In practical and statistical terms only the response distributions from Education specialists show any real points of difference. It would perhaps be expected for Education academics to see the topic of the questionnaire through a slightly different lens, but the results are by no means conclusive.

This chapter has attempted to reveal disciplinary differences in academic perceptions towards the NSS and has described a mixed picture. It can be a reasonably concluded that there are differences between disciplines in general terms and these may reveal themselves more readily in studies exploring other topics. The hint at differences in the perspectives between Education academics and the rest of the respondents suggests this. They could perhaps be described as an “affinity” subject as Braxton (1995) defined them. This specific point would require further investigation, particularly as Braxton’s work is now over fifteen years old. However, where the NSS is concerned, the similarities between the subject areas outweigh the differences heavily. If we take into account the analysis of the data at the macro level and the qualitative analysis we are beginning to see a picture emerge showing a consistent and generally sceptical view towards the NSS as a potential enhancement tool within this specific group of respondents. Perhaps we are seeing an example of the problem Becher (1994) described, namely that the generic nature of the intervention (in this case the NSS) has prevented the survey from gaining credibility for use in departmental enhancement work. This could also explain the perceived difference in levels of priority between senior managers, who have an institutional perspective, from staff in departments. These perceptions paradoxically show some similarities between academics of different disciplinary creeds.

5.2. Institutional differences

Although not one of the core research questions of this study it is worth paying attention to the differences between groups of academic staff from different types of institutions. It is a common strategy for research on the higher education sector to attempt to draw distinctions between institutions of different types and geographical locations, for an example see work by the Higher Education Academy (2009). The differences between nations of the UK are also important because of the different higher education systems each of the nations have. With these considerations in mind this section of the chapter will explore the data collected during the questionnaire at an institutional level.

Table 1 in chapter three shows the number of responses from each of the institutions. There were responses from each of the universities which were included in the sample. However, due to the response rates the numbers were considered too small to run an institution by institution analysis and so it was deemed necessary to aggregate the responses by institutional type. This revealed an issue with the sampling as all of the universities were founded before the 1992 restructuring of UK higher education; with six from the Russell Group and a further two from the 1994 Group (see Table 1). This could be attributed to the disciplinary range chosen, which are generally considered to be traditional subjects and therefore more likely to be taught in the older institutions (noting that an institution needed to teach all three subjects to be included in the sample). This means that the conclusions for this study in general cannot be assumed to be applicable to the whole UK higher education sector as there is a wide variety of institutions that have not been considered in this study. Two aggregations were conducted; one grouping Russell Group with other Pre-92 universities and the other grouping English universities with those from other parts of the UK. Of the 324 responses gathered 285 stated their institution’s name. One hundred and eighty-one respondents were from Russell Group universities and 199 were from English institutions.

Russell Group and other Pre-92 universities

Table 18: Comparison of levels of agreement between Russell Group universities and the other Pre-92 institutions in the sample

Question	Russell Group Disagree %	Russell Group Neither agree nor disagree %	Russell Group Agree %	Other Pre-92 Disagree %	Other Pre-92 Neither agree nor disagree %	Other Pre-92 Agree %	Sig
Q1	1.1	9.7	97.2	1.9	2.9	95.2	0.665
Q2	0.6	5.0	94.4	2.9	2.9	94.2	0.198
Q3	49.7	34.4	15.9	45.9	35.7	18.4	0.812
Q4	3.2	23.7	73.1	4.2	27.4	68.4	0.719
Q5	20.0	43.1	36.9	13.9	33.7	52.5	0.044
Q6	18.8	40.0	41.3	15.3	22.4	62.2	0.003
Q7	46.3	32.7	21.0	41.2	25.8	33.0	0.093
Q8	65.2	27.3	7.5	61.9	26.8	11.3	0.566
Q9	35.4	33.5	31.0	33.3	30.2	36.5	0.664
Q10	88.6	10.2	1.2	83.0	12.8	4.3	0.226
Q11	23.9	36.1	40.0	25.5	42.6	31.9	0.422
Q12	30.0	42.0	28.0	26.6	44.7	28.7	0.843
Q13	21.7	43.4	35.0	24.4	36.7	38.9	0.599
Q14	18.0	18.0	64.0	14.3	17.3	68.4	0.703
Q15	10.4	10.4	73.9	6.8	17.0	76.1	0.267
Q16	63.2	22.4	14.4	61.6	23.2	15.2	0.965
Q17	46.6	26.1	27.3	37.5	32.3	30.2	0.343

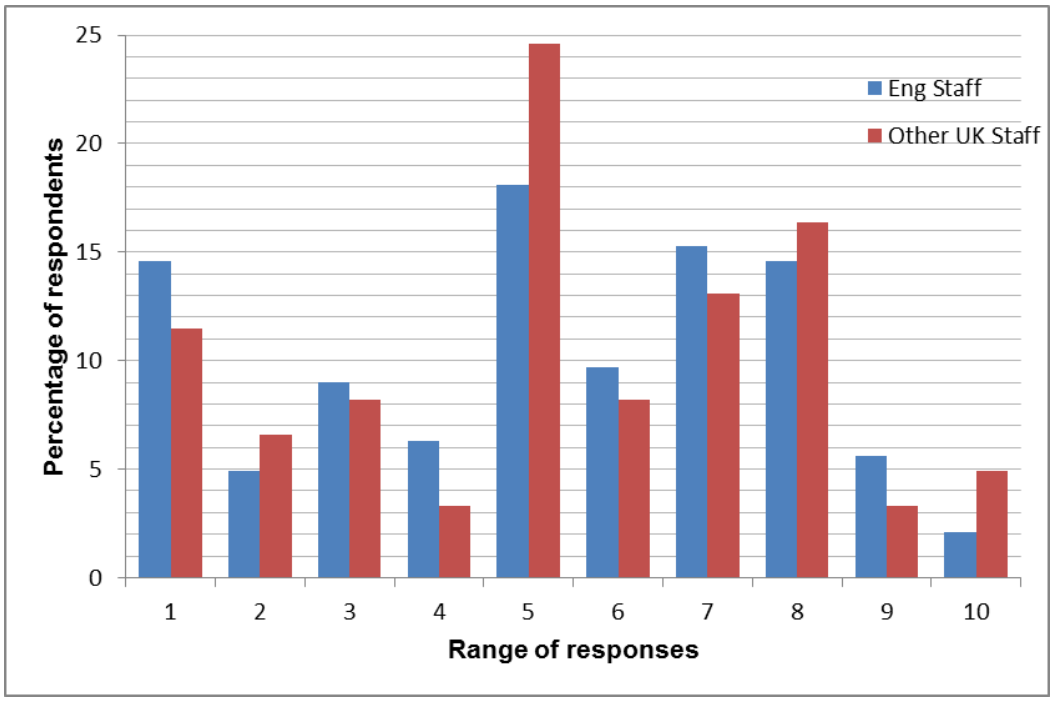
Table 18 shows the levels of agreement with the core 17 items of the questionnaire, disaggregating the responses between Russell Group universities and the other Pre-92 institutions. The two questions showing major differences were the items asking the respondent to rate the extent to which high or low scores in the NSS show something that requires addressing or highlighting. For both of these questions the academic staff from the Russell Group universities agreed less often with these statements. In Q6 "Low scores in the NSS show that there are issues with undergraduate provision that require addressing" the difference in the level of agreement is very large (20.9%). The reasons for these differences are not clear; however perhaps due to the imperative to maintain a very strong reputation, Russell Group university staff are less

likely to see a bad result in the NSS as evidence of poor provision. The reduced, albeit significant difference for Q5 "High scores in the NSS show that there are examples of good practice that might usefully be shared with others" could be explained by the flip side of this point. Perhaps Russell Group university staff are less likely to assume that high scores provide evidence of success. It could be hypothesised that only those institutions performing poorly in league tables would downplay the impact of the NSS as an indicator of problems with provision. However, in general the institutions in the sample were good performers in the league tables, all featuring in the top half through their performance across the range of metrics. We could be seeing evidence of the importance of reputational factors to Russell Group universities which are actually unrelated to league tables or the performance indicators available in higher education. The differences between these universities and others suggest that the difference would be even starker if data was collected from teaching-led institutions, although this was not possible due to the sampling strategy employed.

Nations of the UK

HEFCE have been the driving force behind the development and maintenance of the NSS since its inception. However, Welsh and Northern Irish institutions have always taken part and an increasing number of Scottish institutions are also taking up the survey. This suggests that an interesting analysis might be at the national level, disaggregated between England and the other UK nations.

Figure 19: Ratings of knowledge about issues relating to the NSS disaggregated by nation of the UK



It could be assumed that the English origin of the NSS would lead to academics working in English universities having a deeper knowledge of the survey. There was actually very little difference between those working in English institutions and those in other parts of the UK in terms of their perceived and self-rated knowledge about the NSS. The mean for the other UK staff was 5.36 compared with 5.19 for the England based staff and the standard deviations were very close indeed (2.576 for England based and 2.569 for the other UK staff). Unsurprisingly given the closeness of these distributions the chi-squared test showed that the differences were not statistically significant with $p>0.900$. This would contribute to a rejection of the hypothesis that staff within English institutions are more confident in their knowledge about the NSS and how it can be used. There is no perceived lag in the levels of knowledge arising out of the voluntary uptake of the NSS within non-English institutions.

Table 20 shows the level of agreement when the responses were disaggregated by geographical location of the institutions. The

distributions of responses for each of the core 17 items of the questionnaire were largely similar with a small number of exceptions. Question 4 revealed an interesting difference that is not easy to explain, with the staff based in England agreeing to a greater extent than the other UK academic staff that there are other more appropriate tools than the NSS for gathering views from students. The other UK academic staff placed their responses in the middle of this Likert scale more often. This could be due to the fact that Scottish institutions have chosen to opt in to the NSS suggesting that some value is placed in it as an indicator of student views on the quality of the course, but the endorsement is still not particularly strong as over 60% of the other UK academic staff still agreed with the statement.

Table 20: Comparison of levels of agreement between English universities and institutions from the other parts of the UK

Question	England Disagree %	England Neither agree nor disagree %	England Agree %	Other UK Disagree %	Other UK Neither agree nor disagree %	Other UK Agree %	Sig
Q1	1.5	2.5	96.0	1.2	1.2	97.7	0.754
Q2	1.0	4.5	94.4	2.3	3.5	94.2	0.639
Q3	50.6	34.7	14.7	43.0	35.4	21.5	0.345
Q4	4.0	21.0	75.0	2.7	34.7	62.7	0.072
Q5	18.4	40.8	40.8	15.9	36.6	47.6	0.587
Q6	17.9	36.9	45.3	16.5	25.3	58.2	0.127
Q7	47.0	29.3	23.8	38.5	32.1	29.5	0.421
Q8	66.1	25.0	8.9	59.0	32.1	9.0	0.489
Q9	36.0	33.1	30.9	31.6	30.3	38.2	0.527
Q10	87.2	10.6	2.2	85.2	12.3	2.5	0.910
Q11	23.0	38.5	38.5	28.0	38.7	33.3	0.632
Q12	31.5	43.5	25.0	22.4	42.1	35.5	0.165
Q13	27.2	43.2	29.6	12.7	35.2	52.1	0.002
Q14	16.7	19.4	63.9	16.5	13.9	69.6	0.544
Q15	10.5	11.8	77.8	5.7	15.7	78.6	0.410
Q16	60.2	22.5	17.3	68.3	23.2	8.5	0.165
Q17	44.7	30.2	25.1	39.7	24.4	35.9	0.205

Question 13 asked respondents to rate their agreement with the statement "My institution could use the NSS data more effectively than it

currently does". The difference between the distributions of responses is marked and highly statistically significant. Less than 30% of those in England agreed with the idea that their institution could use the NSS more effectively while over 50% of the other UK academics felt this was the case. This leads to two possible conclusions; either the English based academic staff feel their institutions are performing better than their non-English counterparts when it comes to using the NSS data or the staff in England see less potential in using the NSS, believing their institution could not do any better and the survey has reached its limit of usefulness. The distributions in responses for the other items of the questionnaire begin to give some clues as to which interpretation is more appropriate. As highlighted above, there was a large difference in Q4 with English based staff believing to a greater extent that the NSS is not the best way of gathering student feedback. Q6 showed an important difference also, with a higher percentage of other UK academic staff agreeing that low NSS scores show something that requires addressing. Q17 also reveals a clue as over 10% more of other UK academic staff agreed that the NSS is a useful tool for improving teaching, although this was still just over a third of staff. These responses suggest that the latter interpretation is more appropriate. It seems as if English based academic staff believe the NSS has reached a limit of usefulness in its current form and that other UK academic staff believe the NSS to still have untapped potential, which is why they felt their institutions could use the NSS more effectively when asked during the questionnaire.

This part of the study has revealed some interesting differences between institutions of different types and different locations. There is an indication that staff from non-English institutions felt slightly more positively about the NSS than their England based counterparts although they cannot be regarded as having had a positive perspective on the whole. Russell Group academics made up a large proportion of the sample and on the whole their results are not dissimilar from the rest of the respondents. However there is some indication that when it is used to make sweeping statements about quality, the Russell Group academics are less likely to

agree with the conclusions of the NSS. This could be a rejection of the contribution NSS can make to discussions of institutional reputation, and this is especially so when the scores are low.

Chapter five has developed analyses of the data collected at both the disciplinary and the institutional level and has unearthed some interesting minutiae within the data. In the next chapter these will be combined with the overall findings to help conclude the study and establish where the research in this area may head next.

6. Conclusions

This study has gathered quantitative and qualitative data from over 300 academic staff each offering their perspective on the NSS and its potential use as a tool for enhancement. This concluding chapter will determine the extent to which this study has responded to the research questions specified at the outset. It will then go further by suggesting implications for policy and practice and directions for further research that build upon this study.

6.1. Research questions of this study

In the first chapter a number of research questions were specified. The first of these was whether or not academic staff felt that the NSS was a reliable indicator of teaching quality. Several items within the questionnaire have contributed to the development of an answer to this question giving an indication of the feelings of staff on this issue. For example, when asked to say whether they felt the NSS is a suitable measure of teaching quality, only a small proportion agreed with the statement. There were also low levels of agreement with the idea that low and/or high scores were indicative of good practice or problems which needed addressing. This particular scepticism was more prevalent in the Russell Group universities. The number of people disagreeing with the statement that the NSS is a useful tool for improving teaching far outnumbered the number who agreed with it. If usefulness of a survey tool is assumed to be linked to its perceived reliability, this would suggest that the majority of academic staff in these institutions do not see the NSS as a reliable indicator of teaching quality. This could be a result of the context in which the NSS works and the reaction this causes within the academic community. As was seen throughout the section looking at how the NSS is used, it appears to be very much a top-down initiative and this can cause a reaction that is actually unwarranted as there is some evidence to suggest that the NSS is both valid and reliable at an

institutional level and when comparing within subject areas. The rejection of the NSS's reliability by the academic community may not be a result of the statistical merit of the instrument itself.

The next two research questions were closely linked, with one asking whether or not academic staff use the NSS for enhancement and the other asking about other ways the data is used within institutions. It is clear from the qualitative data highlighted in chapter four that the NSS is used in a variety of different ways and some of these are enhancement led. Senior managers were seen as the major drivers behind the requirement to use the NSS data and this also came through within the open comments. This intervention was not always seen as helpful with academic staff often highlighting at the same time the impact of league tables and marketing to the institution they serve. As a result the NSS was shown to be a feature of the quality systems set up by universities for use across the institution. At the departmental level this would involve discussing the results within teaching and learning committees. Individual academic staff appeared willing to make the adjustments suggested by their department. There were some members of staff who suggested that they would reflect upon the NSS results themselves, but the view in general was that the NSS is actually one step removed from their own teaching. Respondents referred to the relevance of their end of module surveys or the way they engage with students to gather feedback. Some staff went further by suggesting that the NSS was a distraction, preventing them from using more helpful strategies to inform enhancement work. These points were also supported in the quantitative part of the questionnaire. During the development of the NSS it was shown that staff believed a national level survey would add little to the departmental mechanisms already in place (CHERI, 2003). This feeling does seem to be prevalent several years later. As the NSS was set up to provide a performance indicator and public information, it is proving difficult for academic staff to see the value in using the generic instrument for their specific purposes. The top-down nature of the survey, from government through senior management exacerbates this issue as does

the influence of league tables. League tables are particularly interesting as they appear to be encouraging the chasing of raw scores at the expense of meaningful critiques of the available data. As it currently stands these two original purposes of the NSS are incompatible with the third as a tool for enhancement. Although these three purposes were seen as coexisting by CHES (2010), this cannot be achieved without a great deal more understanding within the sector about what the NSS shows and how the available data can be usefully interpreted. There were some respondents to the questionnaire who used the NSS as one of several diagnostic tools to establish the strengths and weaknesses of their provision. Potentially, this could provide a useful route for determining areas in which to target enhancement activities.

The literature review explored disciplinary differences within higher education and one of the research questions asked whether or not these disciplinary differences were reflected in views on the NSS. Chapter five explored this in detail and found a remarkable similarity at the macro level in the distribution of results for each of the three disciplines studied in this research. There were no major deviations in the means or standard deviations and only a few statistically significant differences were found when each subject was compared against the rest of the sample. Education academic staff appeared to have a slightly different perspective on the way that the survey could be utilised within their department. The tentative conclusion is that these differences could be evidence of Education as an “affinity” subject (Braxton, 1995) but this is by no means certain. In total, the similarities outweigh the differences indicating an overarching perspective on the NSS which is not determined or affected by disciplinary background. The generic nature of the NSS is likely to have caused the more general rejection of the survey, with doubts being expressed about its usefulness and relevance. If the primary identity of an academic is related to their discipline and generic tools are therefore less well regarded (Becher, 1994) the NSS is likely to be suffering this fate in many cases.

Smeby's (1996) conclusion that disciplinary norms and institutional context intertwine when developing content actually leads to a double effect in the case of the NSS. There is a double rejection of the NSS due to the views of staff through a disciplinary lens as well as through an institutional one. The former is due to the generic nature of the NSS as an intervention in higher education and the latter is because of institutional priorities which have arisen through the non-enhancement purposes of the NSS. This context has actually damaged the ability of the NSS to be effectively used for enhancement, despite the merits it has when used with other forms of data. This point is of crucial importance not just for the use of student surveys but for the more general design of interventions aimed at improving teaching and learning. If the intervention does not have the ownership of those with the closest connection to the students i.e. academic staff, it is far less likely to have the necessary buy-in required to make it successful.

6.2. Implications of this study

This research had shed additional light on the views often expressed in the media and in policy discussions relating to the NSS. As suggested in the introduction there continues to be a number of different perspectives and requirements from this national level survey. There is evidence contained within this study showing possibilities for using the NSS in a meaningful way for enhancement purposes. However because of the context in which the NSS functions it appears to be difficult to separate the NSS's original purposes as a performance indicator and source of public information from the third purpose it was given later. In fact, these seem to be incompatible in many ways. The top-level information provided for simple public consumption is not the data which departments find useful when developing enhancement activities. This is shown by the preference found during this study for tailored end of module/course questionnaires and other forms of student feedback. There are implications for policy at three different levels: departmental; institutional and national, each of which have a contribution to make in ensuring that

the NSS can be used more effectively to develop enhancement activities. At the departmental level there often appears to be a disconnect between the NSS and other forms of student feedback. For this reason the NSS is an add-on with which not all academics engage and its relevance to the individual member of staff is not always clear. This connection needs to be more clearly articulated and information needs to be provided about how the NSS can be used for comparative purposes or for diagnosing issues with provision. This activity can be supported at the institutional level by the development of supportive structures that take the discussion away from league table position and towards enhancement work. The perceived stick wielding of senior managers was seen by many staff as unhelpful. A shift in emphasis towards enhancement would potentially lead to improved student experiences and therefore higher NSS scores.

This study has provided evidence which is at odds with some of the findings of the report *Enhancing and Developing the National Student Survey* (CHES, 2010). This report gathered the views of a small number of sector specialists and stated that the NSS was accepted across the sector. The report also stated that people were relaxed about the three current purposes of the NSS. On both of these points this study has found fresh evidence. There appears to be a lot of work still to be done to show the value of the NSS to the wider academic community as a tool for the improvement of teaching. There is also a clear tension between the two original purposes of the NSS and the third as a result of well embedded policy drivers within universities. There is potential for this to become more pronounced as the higher education market becomes more competitive. More information needs to be provided to the sector about the uses and abuses of NSS data to unlock its enhancement potential to a greater degree.

6.3. Directions for future study

This research has used a mixed methods approach to analyse the views gathered from a wide a range of academic staff at Pre-92 universities

within a pre-defined set of disciplinary areas. As highlighted in chapter three there are a number of potential biases in this study and an important direction for further research would be to explore these biases in order to build a fuller picture of the perceptions of academic staff beyond the current sample. There are some elements of the study that could have been revised in order to make the results more conclusive. For example the choice of disciplines, selected out of personal interest, adversely affected the range of institutions by inadvertently ensuring each of the chosen institutions were Pre-92 universities. The interesting differences between Russell Group universities and the rest of the sample showed the potential for future analyses across a wider range of institutions. A key way of expanding this research would be to use this questionnaire tool across a wider institutional sample, to incorporate the views of Post-92 universities and small and specialist colleges. It might be useful to extend this into Further Education Colleges because of the differences in the types of higher education they offer. This expansion of the sample is particularly important because of the widely assumed differences between institutions of certain types. An example would be that staff within teaching-led institutions are more interested in improving their teaching skills. This would therefore imply a greater affinity with using data provided by students to improve provision. Whether or not this is actually the case is not currently known because of the construction of the sample in this study. It would also be advisable to widen the range of disciplines covered in future studies, as a minimum incorporating a fourth subject from the “hard-applied” part of Biglan’s (1973) typology. Only when these areas are further explored can a more generally applicable conclusion be drawn for the sector as a whole, rather than just the sample surveyed here.

Another factor that may affect the perceptions of academic staff towards the National Student Survey is the position an individual holds within the department or institution. As highlighted above, the relationship with senior management is a theme emerging from this study and so knowing the difference between the views of those in senior positions compared

with those more junior could be an interesting direction for further research. It is posited that the cultural and experiential perspectives of senior staff are different to those of junior colleagues. Although data relevant to this was captured through the questionnaire, the quality and consistency of these data was not high enough to perform a reliable analysis, partly because the questionnaire asked staff to state their response in an open comment. A more appropriate way of capturing this information would be to ask respondents to choose from a comprehensive closed list of job titles to improve the quality of the dataset.

The questionnaire used within this study was piloted effectively and utilised the expertise of a number of colleagues with credentials in survey design. The questionnaire tool as it stands is a useful way of gathering data about the perceptions of academic staff towards the NSS. However in its current form the questionnaire does not allow the connection of perception with some of the realities of that institutional context. In other words it was the perspective of the “average academic” that this questionnaire was designed to explore. It would be an interesting direction of future research to explore in an in depth way the methods that are used within universities to utilise the survey data. A case study approach within a small number of institutions could be enlightening. In particular this could unpick the issues highlighted above about the role of senior management within universities. This study has not had opportunity to explore the motivations and aspirations which have influenced their strategies relating to the NSS. This would be a very interesting perspective indeed.

As a final conclusion, any student survey conducted at a national level must overcome a number of hurdles before it becomes embedded as a way of providing data about student experiences at university. This study has highlighted a number of those barriers. The NSS has overcome several of these, namely the requirement for political will to provide information to students about university life and an increasing expectation

that students will be included as part of the discussion about enhancement initiatives. However, the NSS has still not won the hearts and minds of the staff who are charged with reacting to the results and this prevents a lot of useful activity from taking place. There are significant doubts within the current academic community about the usefulness of the NSS and this cannot be ignored by the policymakers who commission the survey.

(Word count: 28333)

7. References

Attwood, R. (2010, September 30). Hollow victory: 12 Angry Scholars Berate 'Risible' NSS. *Times Higher Education,* Available from http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=413687 [accessed 16 October 2010]

Ballantyne, R., Borthwick, J. and Packer,J. (2000). Beyond student evaluations of teaching: identifying and addressing academic staff development needs. *Assessment and Evaluation in Higher Education, 25*, 221-236.

Barnett, V. (1991). *Sample survey principles and methods*. London: Edward Arnold.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*, 151-61.

Biglan, A. (1973). The characteristics of subject matter in different scientific areas. *Journal of Applied Psychology, 57*, 195-203.

Braxton, J. (1995). Disciplines with an Affinity for the Improvement of Undergraduate Education, in Hatvia, N. and Marincovich, M. (Eds). *Disciplinary Differences in Teaching and Learning: Implications for Practice*. San Francisco: Jossey-Bass.

Broomfield, D., and Bligh, J. (1998). An evaluation of the 'short form' course experience questionnaire with medical students. *Medical Education, 32*, 367-369.

Cambridge University Students' Union. (2010). *The National Student Survey*. Retrieved 16 October, 2010 from http://www.cusu.cam.ac.uk/campaigns/education/nss/

Cashin, W. E., and Downey, R. G. (1992). Using Global Student Rating Items for Summative Evaluation, *Journal of Educational Psychology, 84* (4), 563–72.

Centre for Higher Education Studies. (2010). *Enhancing and Developing the National Student Survey*. Bristol: HEFCE.

Chen, C.K., and Hughes, J. (2004). *Using Ordinal Regression Model to Analyze Student Satisfaction Questionnaires*. Tallahassee: Association for Institutional Research

CHERI, NOP Research Group., and SQW Ltd. (2003). *Collecting and using student feedback on quality and standards of learning and teaching in HE*. Bristol: HEFCE.

CHERI., Open University., and Hobsons Research. (2008). *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England*. Bristol: HEFCE.

Cohen, P. (1980). Effectiveness of student-rating feedback for improving college instruction: A Meta-Analysis of Findings. *Research in Higher Education, 13*, 321-341.

Cox, E.P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research, 17* (4), 407-422.

Creswell, J.W. (2003). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. London: SAGE.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3), 297-334.

d'Apollonia, S., and Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.

Davis, L. (1992). Instrument Review: Getting the Most From a Panel of Experts. *Applied Nursing Research, 5*. 194-197.

Dent, P., and Nicholas, T. (1980). A Study of Faculty and Student Opinions on Teaching Effectiveness Ratings. *Peabody Journal of Education, 57*, 135-147.

De Vaus, D. (2002). *Surveys in Social Research 5th Edition*. London: Routledge.

Dillman, D., Smyth, J., and Melani Christian, L. (2009). *Internet, Mail and Mixed-mode Surveys: The tailored design method 3rd edition*. New Jersey: John Wiley & Sons.

Entwistle, N., and Tait, H. (1995). Approaches to Studying and Perceptions of the Learning Environment Across Disciplines, in Hatvia, N. and Marincovich, M. (Eds). *Disciplinary Differences in Teaching and Learning: Implications for Practice*. San Francisco: Jossey-Bass.

Enyon, R. (2005). The use of the internet in higher education: Academics' experiences of using ICTs for teaching and learning. *Aslib Proceedings: New Information Perspectives, 57*, 168-180.

Falk, B., and Dow, L.D. (1971). *The Assessment of University Teaching*. London: Society for Research into Higher Education.

Flint, A., Oxley, A., Helm, P., and Bradley, S. (2009). Preparing for success: one institution's aspirational and student focused response to the National Student Survey. *Teaching in Higher Education, 14*, 607-618.

Flood Page, C. (1974). *Student Evaluation of Teaching: The American Experience*. London: Society for Research in to Higher Education.

Gibbs, G. (2000). Are the pedagogies of the disciplines really different?, in Rust, D. (Ed). *Improving Student Learning Through the Discipline*. Oxford: Oxford Centre for Staff and Learning Development.

Glaser, B.G., and Strauss A.L. (1967). *The Discovery of Grounded Theory*. New York: Aldine de Gruyter.

Greene, J.C., Caracelli, V.J., and Graham, W.F. (1989). Towards a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11* (3), 255-274.

Gregory, R., Harland, G., and Thorley, L. (1995). Using a student experience questionnaire for improving teaching and learning, in Gibbs, G. (Ed.) *Improving student learning through assessment and evaluation*. Oxford: Oxford Centre for Staff Development.

Hanbury, A. (2007). *Comparative Review of British, American and Australian national surveys of undergraduate students*. York: Higher Education Academy.

Harvey, L. (1997). *Student satisfaction manual*. Buckingham: Society for Research into Higher Education.

Harvey, L. (2001). *Student Feedback: a report to the Higher Education Funding Council for England*. London: HEFCE.

Harvey, L. (2008, June 12). Jumping through hoops on a white elephant: a survey signifying nothing. *Times Higher Education,* Available online at http://www.timeshighereducation.co.uk/story.asp?storyCode=402335&sectioncode=26 [accessed 16 October 2010]

Hatvia, N. and Birenbaum, M. (2000). Who prefers what? Disciplinary Differences in Students' Preferred Approaches to Teaching and Learning Styles. *Research in Higher Education, 41*, 209-236.

HEFCE. (2002). *Information on quality and standards in higher education*. Bristol: HEFCE.

HEFCE. (2004). *National Student Survey 2005: Outcomes of consultation and guidance on next steps*. Bristol: HEFCE.

HESA. (2012). *Staff in Higher Education Institutions*. Cheltenham: HESA.

Higher Education Academy. (2009). *Reward and Recognition of teaching in Higher Education*. York: Higher Education Academy.

Hudson, D., Seah, H-L., Hite, D. and Haab, T. (2004). Telephone presurveys, self-selection, and non-response bias to mail and Internet surveys in economic research, *Applied Economics Letters,* 11 (4), 237-240.

Johnson, R.B., and Onwuegbuzie, A.J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33* (7), 14-26.

Kember, D., Leung D.Y.P., and Kwan, K.P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*, 411-425.

Kerlinger, F.N. (1971). Student evaluation of university professors. *School and society, 99*, 353-356.

Kolditch, E., and Dean, A.V. (1999). Student Ratings of Instruction in the USA: hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education, 24*, 27-42.

Lyon, P.M., and Hendry, G.D. (2002). The Use of the Course Experience Questionnaire as a Monitoring Evaluation Tool in a Problem Based Medical Programme. *Assessment and Evaluation in Higher Education,* 339-352.

Marsh, C. (1982). *The Survey Method: The Contribution of Surveys to Sociological Explanation*. London: George Allen and Unwin.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research*. International Journal of Educational Research, 11*, 253-388.

Marsh, H., and Cheng, H. (2008). *National Student Survey of Teaching in UK Universities: Dimensionality, Multilevel Structure and Differentiation at the Level of University and Discipline: Preliminary Results*. York: Higher Education Academy.

Marsh, H., and Roche, L. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal, 30*, 217-251.

Moore, S., and Kuol, N. (2005). Students evaluating teachers: exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education, 10*, 57-73.

Mostrous, A. (2008, May 14). Kingston University students told to lie to boost college's rank in government poll. *The Times,* Available at http://www.timesonline.co.uk/tol/news/uk/article3924417.ece [accessed 16 October 2010]

Murray, H.G. (1997). Does evaluation of teaching lead to improvement in teaching? *International Journal for Academic Development,2* (1), 8-24.

Muircheartaigh, C., Krosnik, J., and Helic, A. (2000). *Middle Alternatives, Acquiescence, and the Quality of Questionnaire Data*. Chicago: University of Chicago.

Nasser, F., and Fresko, B. (2002). Faculty views on student evaluation of college teaching. *Assessment and Evaluation in Higher Education, 27*, 187-198.

Neumann, R. (2001). Disciplinary Differences and University Teaching. *Studies in Higher Education, 26*, 136-146.

Neumann, R., Parry, S., and Becher, T. (2002) Teaching and Learning in their Disciplinary Contexts: a conceptual analysis. *Studies In Higher Education, 27* (4), 405-418.

Newman, I., and McNeil. K. (1998). *Conducting Survey Research in the Social Sciences*. Lanham: University Press of America.

Oakleigh Consulting., and Staffordshire University. (2010). *Understanding the information needs of users of public information about higher education*. Bristol: HEFCE.

Paulhaus, D.L. (1991). Measurement and control of response bias, in Robinson, J.P., Shaver, P.R. and Wrightsman, L.S. (Eds.), *Measures of personality and social psychological attitudes*. San Diego: Academic Press.

Preston, C.P. and Colman, A.M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power and respondent preferences. *Acta Psychologica, 104*, 1-15.

Ramsden, P., and Entwistle, N.J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology, 51*, 368-383.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*, 129-150.

Richardson, J. T. E. (1994). A British Evaluation of the Course Experience Questionnaire, *Studies in Higher Education,* 19, 59–68.

Richardson, J.T.E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment and Evaluation in Higher Education, 30*, 387-415.

Richardson, J.T.E., Slater, J.B., and Wilson, J. (2007). The National Student Survey: development, findings and implications. *Studies in Higher Education, 32* (5), 557-580.

Rust, D. (2000). *Improving Student Learning Through the Discipline*. Oxford: Oxford Centre for Staff and Learning Development.

Saroyan, A., and Amundsen, C. (2001). Evaluating university teaching: time to take stock. *Assessment and Evaluation in Higher Education, 26*, 341-353.

Schemlkin, L.P., Spencer, K.J., and Gillman, E.S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education, 38*, 575-592.

Sheehan, K. (2001). Email Survey Response Rates: A Review. *Journal of Computer-Mediated Communication, 6* (2), 1-19.

Sheikh, K., and Mattingly, S. (1981). Investigating non-response bias in mail surveys. *Journal of Epidemiology and Community Health, 35*, 293-296.

Smart, J., and Ethington, C. (1995). Disciplinary and Institutional Differences in Undergraduate Education Goals, in Hatvia, N. and Marincovich, M. (Eds). *Disciplinary Differences in Teaching and Learning: Implications for Practice*. San Francisco: Jossey-Bass.

Smeby, J. (1996). Disciplinary differences in university teaching. *Studies in Higher Education, 21*, 69-79.

SPSS. (2002). *Ordinal Regression Analysis, SPSS Advanced Models 10.0*. Chicago: SPSS.

Surridge, P. (2008). *The National Student Survey 2005-07: Findings and Trends*. Bristol: HEFCE.

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

Timpson, W., and Desley, A. (1997). Rethinking student evaluations and the improvement of teaching: Instruments for change at the University of Queensland. *Studies in Higher Education, 22*, 55-65.

Veall, M., and Zimmermann, K. (1996). *Pseudo R^2 Measures for Some Common Limited Dependent Variable Models*. London: CEPR.

Wason, P.C., and Jones, S. (1963). Negatives: Denotation and Connotation. *British Journal of Psychology, 54* (4), 299-307.

Weems, G H., Onwuegbuzie, A. J., Schreiber, J. B. and Eggers, S. J. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales, *Assessment & Evaluation in Higher Education*, 28, 587-606.

Wembridge, E.R., and Means, E.R. (1918). Obscurities in voting upon measures due to double-negative. *Journal of Applied Psychology, 2*, 156-63.

Wiers-Jenssen, J., Stensaker, B., and Grogaard, J.B. (2002). Student satisfaction: Towards an empirical deconstruction of the concept. *Quality in Higher Education, 8*, 183-195.

Williams, J., and Kane, D. (2008). *Exploring the National Student Survey: Assessment and feedback issues*. York: Higher Education Academy.

Wilson, K., Lizzio, A., and Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education, 22*, 33-53.

Winship, C., and Mare, R.D. (1992). Models for Sample Selection Bias. *Annual Review of Sociology, 18*, 327-350.

Yorke, M. (1995). Taking the Odds-On Chance: Using Performance Indicators in Managing for the Improvement of Quality in Higher Education. *Tertiary Education and Management, 1*, 49-57.

Yorke, M. (2009). Student experience' surveys: some methodological considerations and an empirical investigation. *Assessment & Evaluation in Higher Education, 34*, 721-739.

8. Appendices

Appendix 1: Email text used for survey distribution

Dear [name],

I am writing to ask for your participation in a survey I am conducting as part of my research Masters at the University of York. I am asking academics about their perceptions towards the National Student Survey (NSS) as a tool for the improvement of teaching within universities.

This is a short survey and should take no longer than 10 minutes to complete. Please follow the link below to go to the survey website.

[Hyperlink to survey]

Your responses will be kept anonymous. Should you have any further queries about my research please feel free to email me at ac516@york.ac.uk or phone me using 07921164155.

Thank you very much for your time in completing this survey.

Yours sincerely,
Adam

Appendix 2: The questionnaire used in this study

Please rate your own knowledge of issues relating to the National Student Survey (NSS) 1= low levels of knowledge 10= high levels of knowledge

To what extent to you agree with the following statements?

1. Knowing what students think about their course is important when seeking to improve teaching
2. Students should have an opportunity to rate the quality of their course
3. The NSS is a suitable measure of teaching quality
4. There are other more appropriate tools than the NSS for gathering views from students about the quality of their course
5. High scores in the NSS show that there are examples of good practice that might usefully be shared with others
6. Low scores in the NSS show that there are issues with undergraduate provision that require addressing
7. The NSS provides useful information to help me improve my teaching
8. I think that my own teaching has improved as a result of making changes informed by NSS data
9. The NSS results tell me information that I would like to know

10. The NSS is my preferred method for gathering feedback from students
11. The NSS distracts colleagues from other possible ways to improve teaching and learning
12. My department/faculty could use the NSS data more effectively than it currently does
13. My institution could use the NSS data more effectively than it currently does
14. The NSS is generally more of a concern to senior management than to teachers within departments/faculties
15. My institution shares results with individual departments/faculties
16. League tables are a positive development in Higher Education
17. Overall, I see the NSS as being a useful tool for improving teaching in Higher Education

In your working environment where does the request to respond to the NSS results usually come from? Select all that apply

- Senior management within the institution
- Colleagues within my department
- From myself as an individual member of staff
- From students
- There is no requirement to use NSS data
- Other (please specify)

How does your department or faculty use the results of the NSS? Why?

How do you, as an individual, use the results of the NSS? Why?
Please select your gender

- Male
- Female

Please state your current job title within your university

This study is focusing on academics from three subject areas: Education, History and Physics. Which of these subject areas do you belong to?

- Education
- History
- Physics
- Other (please specify)

Follow up interviews with a small number of academic staff will be conducted following the closure of this survey. Please enter your email address if you would you be willing to participate in one of these interviews?

Appendix 3: Full data for distribution of responses by gender

Question	Male n	Male Disagree %	Male Neither agree nor disagree %	Male Agree %	Male mean response	Male standard deviation	Female n	Female Disagree %	Female Neither agree nor disagree %	Female Agree %	Female mean response	Female standard deviation	Sig
Q1	201	1.0	3.5	95.5	2.95	0.27	108	1.9	0.9	97.2	2.95	0.29	0.334
Q2	201	1.0	4.0	95.0	2.94	0.28	107	1.9	4.7	93.5	2.92	0.34	0.776
Q3	167	50.9	30.5	18.6	1.68	0.77	97	47.4	41.2	11.3	1.64	0.68	0.123
Q4	173	5.8	24.3	69.9	2.64	0.59	96	0.0	25.0	75.0	2.75	0.44	0.056
Q5	176	18.2	38.6	43.2	2.25	0.74	103	19.4	40.8	39.8	2.20	0.75	0.858
Q6	174	16.1	31.6	52.3	2.36	0.75	101	19.8	38.6	41.6	2.22	0.76	0.230
Q7	182	45.1	28.6	26.4	1.81	0.83	96	47.9	30.2	21.9	1.74	0.80	0.711
Q8	181	65.2	26.0	8.8	1.44	0.65	98	64.3	27.6	8.2	1.44	0.64	0.950
Q9	178	32.0	34.8	33.1	2.01	0.81	96	40.6	28.1	31.3	1.91	0.85	0.325
Q10	185	86.5	10.8	2.7	1.16	0.44	93	87.1	11.8	1.1	1.14	0.38	0.664
Q11	174	27.6	33.3	39.1	2.11	0.81	92	15.2	50.0	34.8	2.20	0.68	0.014
Q12	164	27.4	46.3	26.2	1.99	0.73	96	32.3	39.6	28.1	1.96	0.78	0.549
Q13	159	21.4	44.7	34.0	2.13	0.74	90	26.7	37.8	35.6	2.09	0.79	0.503
Q14	179	15.1	16.8	68.2	2.53	0.74	98	19.4	19.4	61.2	2.42	0.80	0.492
Q15	155	10.3	11.6	78.1	2.68	0.65	86	8.1	16.3	75.6	2.67	0.62	0.541
Q16	193	63.7	21.2	15.0	1.51	0.74	101	62.4	25.7	11.9	1.50	0.70	0.583
Q17	178	43.8	28.1	28.1	1.84	0.84	98	45.9	28.6	25.5	1.80	0.82	0.895

Appendix 4: Full data for distribution of responses for lower and higher perceived levels of knowledge about the NSS

Question	1-5 n	1-5 Disagree %	1-5 Neither agree nor disagree %	1-5 Agree %	1-5 mean response	1-5 standard deviation	6-10 n	6-10 Disagree %	6-10 Neither agree nor disagree %	6-10 Agree %	6-10 mean response	6-10 standard deviation	Sig
Q1	124	1.6	4.0	94.4	2.93	0.32	109	0.9	1.8	97.2	2.96	0.23	0.549
Q2	123	1.6	4.1	94.3	2.93	0.32	109	0.9	3.7	95.4	2.94	0.27	0.880
Q3	87	44.8	46.0	9.2	1.64	0.65	108	55.6	25.9	18.5	1.63	0.78	0.008
Q4	92	1.1	34.8	64.1	2.63	0.51	105	6.7	20.0	73.3	2.67	0.60	0.015
Q5	98	17.3	39.8	42.9	2.26	0.74	109	19.3	42.2	38.5	2.19	0.74	0.813
Q6	94	17.0	35.1	47.9	2.31	0.75	109	22.0	35.8	42.2	2.20	0.78	0.604
Q7	97	48.5	32.0	19.6	1.71	0.78	107	48.6	28.0	23.4	1.75	0.81	0.741
Q8	102	73.5	21.6	4.9	1.31	0.56	104	59.6	29.8	10.6	1.51	0.68	0.082
Q9	95	35.8	37.9	26.3	1.91	0.79	107	31.8	33.6	34.6	2.03	0.82	0.446
Q10	101	86.1	11.9	2.0	1.16	0.42	104	89.4	10.6	0.0	1.11	0.31	0.333
Q11	93	23.7	48.4	28.0	2.04	0.72	106	20.8	32.1	47.2	2.26	0.78	0.016
Q12	84	28.6	42.9	28.6	2.00	0.76	106	35.8	43.4	20.8	1.85	0.74	0.378
Q13	78	21.8	46.2	32.1	2.10	0.73	104	28.8	41.3	29.8	2.01	0.77	0.558
Q14	98	17.3	23.5	59.2	2.42	0.77	107	18.7	16.8	64.5	2.46	0.79	0.493
Q15	86	16.3	16.3	67.4	2.51	0.76	97	2.1	9.3	88.7	2.87	0.40	0.001
Q16	115	61.7	24.3	13.9	1.52	0.73	108	63.0	22.2	14.8	1.52	0.74	0.926
Q17	97	40.2	37.1	22.7	1.82	0.78	106	44.3	24.5	31.1	1.87	0.86	0.125

Appendix 5: Correlation matrix of core questionnaire items

Questions	Know.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17
Know.	1	0.01	-0.03	-0.07	0.08	-0.04	-0.10	0.00	0.13	0.10	-0.03	*0.16*	*-0.18*	-0.12	0.09	**0.32**	-0.04	0.01
Q1		1	**0.54**	**0.20**	0.00	**0.19**	**0.26**	**0.20**	**0.19**	**0.16**	0.00	**-0.18**	0.00	-0.05	**-0.22**	0.11	**0.16**	**0.22**
Q2			1	**0.24**	-0.04	**0.29**	**0.24**	**0.27**	**0.31**	**0.24**	0.10	**-0.28**	-0.01	-0.02	**-0.20**	**0.19**	**0.18**	**0.27**
Q3				1	**-0.42**	**0.56**	**0.45**	**0.54**	**0.56**	**0.49**	**0.47**	**-0.52**	0.08	0.01	**-0.21**	0.03	**0.36**	**0.66**
Q4					1	**-0.27**	**-0.23**	**-0.33**	**-0.36**	**-0.23**	**-0.37**	**0.30**	-0.08	-0.02	*0.15*	0.12	**-0.26**	**-0.38**
Q5						1	**0.60**	**0.54**	**0.47**	**0.50**	**0.31**	**-0.45**	*0.14*	0.07	**-0.22**	-0.06	**0.39**	**0.64**
Q6							1	**0.37**	**0.37**	**0.35**	**0.19**	**-0.38**	**0.17**	0.10	-0.06	0.03	**0.20**	**0.44**
Q7								1	**0.73**	**0.62**	**0.48**	**-0.36**	**0.18**	0.06	**-0.31**	-0.01	**0.32**	**0.65**
Q8									1	**0.56**	**0.60**	**-0.41**	0.10	-0.01	**-0.25**	0.02	**0.32**	**0.63**
Q9										1	**0.38**	**-0.37**	*0.15*	0.07	**-0.23**	0.09	**0.28**	**0.60**
Q10											1	**-0.28**	*0.13*	0.11	*-0.15*	*-0.16*	**0.29**	**0.46**
Q11												1	0.00	0.06	**0.30**	0.09	**-0.34**	**-0.51**
Q12													1	**0.82**	*0.13*	-0.10	0.03	*0.14*
Q13														1	*0.16*	-0.05	-0.03	0.03
Q14															1	0.09	**-0.17**	**-0.21**
Q15																1	-0.04	-0.02
Q16																	1	**0.48**
Q17																		1

Italicised figures indicate correlations are significant at the 0.05 level
Bold figures indicate correlations are significant at the 0.01 level

Appendix 6: Full data for comparison of levels of agreement between Education and the other subjects combined

Question	Education n	Education Disagree %	Education Neither agree nor disagree %	Education Agree %	Education mean response	Education standard deviation	Others n	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Others mean response	Others standard deviation	Sig
Q1	96	2.1	1.0	96.9	2.95	0.30	228	0.9	3.1	96.1	2.95	0.25	0.382
Q2	95	1.1	5.3	93.7	2.93	0.30	228	1.3	3.9	94.7	2.93	0.30	0.855
Q3	78	57.7	34.6	7.7	1.50	0.64	198	47.5	34.3	18.2	1.71	0.76	0.074
Q4	81	0.0	24.7	75.3	2.75	0.43	200	5.5	24.5	70.0	2.65	0.58	0.096
Q5	85	20.0	37.6	42.4	2.22	0.76	206	18.4	40.8	40.8	2.22	0.74	0.878
Q6	84	17.9	35.7	46.4	2.29	0.75	202	17.3	35.1	47.5	2.30	0.75	0.985
Q7	83	49.4	27.7	22.9	1.73	0.81	207	45.9	29.5	24.6	1.79	0.81	0.864
Q8	82	70.7	23.2	6.1	1.35	0.60	209	62.7	28.2	9.1	1.46	0.66	0.409
Q9	82	39.0	31.7	29.3	1.90	0.83	204	34.3	33.3	32.4	1.98	0.82	0.745
Q10	82	87.8	12.2	0.0	1.12	0.33	206	85.9	11.2	2.9	1.17	0.45	0.291
Q11	78	23.1	35.9	41.0	2.18	0.79	199	22.1	41.2	36.7	2.15	0.75	0.705
Q12	77	26.0	36.4	37.7	2.12	0.79	194	31.4	46.4	22.2	1.91	0.73	0.033
Q13	72	25.0	34.7	40.3	2.15	0.80	188	23.4	44.1	32.4	2.09	0.74	0.351
Q14	85	27.1	10.6	62.4	2.35	0.88	203	13.3	20.7	66.0	2.53	0.72	0.006
Q15	73	19.2	6.8	74.0	2.55	0.80	179	5.0	15.6	79.3	2.74	0.54	0.001
Q16	90	71.1	18.9	10.0	1.39	0.67	215	60.0	24.2	15.8	1.56	0.75	0.170
Q17	82	50.0	29.3	20.7	1.71	0.79	204	42.2	28.4	29.4	1.87	0.84	0.292

Appendix 7: Full data for Comparison of levels of agreement between History and the other subjects combined

Question	History n	History Disagree %	History Neither agree nor disagree %	History Agree %	History mean response	History standard deviation	Others n	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Others mean response	Others standard deviation	Sig
Q1	228	0.9	1.7	97.4	2.97	0.22	207	1.4	2.9	95.7	2.94	0.29	0.716
Q2	228	0.9	3.4	95.7	2.95	0.26	206	1.5	4.9	93.7	2.92	0.32	0.739
Q3	198	45.3	34.0	20.8	1.75	0.78	170	53.5	34.7	11.8	1.58	0.69	0.113
Q4	200	6.7	20.0	73.3	2.67	0.60	176	2.3	27.3	70.5	2.68	0.51	0.094
Q5	206	21.3	45.4	33.3	2.12	0.73	183	17.5	36.6	45.9	2.28	0.75	0.109
Q6	202	18.7	36.4	44.9	2.26	0.76	179	16.8	34.6	48.6	2.32	0.75	0.817
Q7	207	43.0	33.6	23.4	1.80	0.79	183	49.2	26.2	24.6	1.75	0.83	0.392
Q8	209	60.4	27.9	11.7	1.51	0.70	180	67.8	26.1	6.1	1.38	0.60	0.194
Q9	204	34.6	29.9	35.5	2.01	0.84	179	36.3	34.6	29.1	1.93	0.81	0.496
Q10	206	83.8	12.4	3.8	1.20	0.49	183	88.0	10.9	1.1	1.13	0.37	0.270
Q11	199	21.4	36.9	41.7	2.20	0.77	174	23.0	41.4	35.6	2.13	0.76	0.594
Q12	194	27.4	46.2	26.4	1.99	0.74	165	31.5	41.8	26.7	1.95	0.76	0.718
Q13	188	19.4	40.8	39.8	2.20	0.75	157	26.8	42.0	31.2	2.04	0.76	0.252
Q14	203	13.1	19.6	67.3	2.54	0.72	181	19.9	16.6	63.5	2.44	0.80	0.318
Q15	179	5.2	15.6	79.2	2.74	0.55	156	11.5	11.5	76.9	2.65	0.68	0.182
Q16	215	63.4	27.7	8.9	1.46	0.66	193	63.2	19.7	17.1	1.54	0.77	0.069
Q17	204	44.3	30.2	25.5	1.81	0.82	180	44.4	27.8	27.8	1.83	0.84	0.875

Appendix 8: Full data for comparison of levels of agreement between Physics and the other subject areas combined

Question	Physics n	Physics Disagree %	Physics Neither agree nor disagree %	Physics Agree %	Physics mean response	Physics standard deviation	Others n	Others Disagree %	Others Neither agree nor disagree %	Others Agree %	Others mean response	Others standard deviation	Sig
Q1	80	0.0	6.3	93.8	2.94	0.24	244	1.6	1.2	97.1	2.95	0.28	0.023
Q2	80	0.0	3.8	96.3	2.96	0.19	243	1.6	4.5	93.8	2.92	0.32	0.487
Q3	67	49.3	32.8	17.9	1.69	0.76	209	50.7	34.9	14.4	1.64	0.72	0.776
Q4	70	5.7	27.1	67.1	2.61	0.60	211	3.3	23.7	73.0	2.70	0.53	0.528
Q5	73	13.7	37.0	49.3	2.36	0.71	218	20.6	40.8	38.5	2.18	0.75	0.208
Q6	72	13.9	34.7	51.4	2.38	0.72	214	18.7	35.5	45.8	2.27	0.76	0.583
Q7	74	50.0	25.7	24.3	1.74	0.83	216	45.8	30.1	24.1	1.78	0.81	0.749
Q8	72	66.7	25.0	8.3	1.42	0.64	219	64.4	27.4	8.2	1.44	0.64	0.923
Q9	71	33.8	35.2	31.0	1.97	0.81	215	36.3	32.1	31.6	1.95	0.82	0.880
Q10	76	88.2	10.5	1.3	1.13	0.38	212	85.8	11.8	2.4	1.17	0.43	0.817
Q11	72	25.0	45.8	29.2	2.04	0.74	205	21.5	37.6	41.0	2.20	0.77	0.204
Q12	64	32.8	50.0	17.2	1.84	0.70	207	29.0	41.5	29.5	2.00	0.77	0.148
Q13	62	27.4	51.6	21.0	1.94	0.70	198	22.7	38.4	38.9	2.16	0.77	0.033
Q14	72	12.5	25.0	62.5	2.50	0.71	216	19.0	15.3	65.7	2.47	0.79	0.118
Q15	60	5.0	20.0	75.0	2.70	0.56	192	10.4	10.9	78.6	2.68	0.65	0.110
Q16	76	59.2	19.7	21.1	1.62	0.82	229	64.6	23.6	11.8	1.47	0.70	0.128
Q17	74	40.5	27.0	32.4	1.92	0.86	212	45.8	29.2	25.0	1.79	0.82	0.459

Appendix 9: Full data for comparison of levels of agreement between Russell Group universities and the other Pre-92 institutions in the sample

Question	Russell Group n	Russell Group Disagree %	Russell Group agree nor disagree %	Russell Group Agree %	Russell Group mean response	Russell Group standard deviation	Other Pre-92 n	Other Pre-92 Disagree %	Other Pre-92 Neither agree nor disagree %	Other Pre-92 Agree %	Other Pre-92 mean response	Other Pre-92 standard deviation	Sig
Q1	181	1.1	9.7	97.2	2.96	0.24	104	1.9	2.9	95.2	2.93	0.32	0.665
Q2	180	0.6	5.0	94.4	2.94	0.26	104	2.9	2.9	94.2	2.91	0.37	0.198
Q3	151	49.7	34.4	15.9	1.66	0.74	98	45.9	35.7	18.4	1.72	0.76	0.812
Q4	156	3.2	23.7	73.1	2.70	0.53	95	4.2	27.4	68.4	2.64	0.56	0.719
Q5	160	20.0	43.1	36.9	2.17	0.74	101	13.9	33.7	52.5	2.39	0.72	0.044
Q6	160	18.8	40.0	41.3	2.23	0.74	98	15.3	22.4	62.2	2.47	0.75	0.003
Q7	162	46.3	32.7	21.0	1.75	0.78	97	41.2	25.8	33.0	1.92	0.86	0.093
Q8	161	65.2	27.3	7.5	1.42	0.63	97	61.9	26.8	11.3	1.49	0.69	0.566
Q9	158	35.4	33.5	31.0	1.96	0.82	96	33.3	30.2	36.5	2.03	0.84	0.664
Q10	166	88.6	10.2	1.2	1.13	0.37	94	83.0	12.8	4.3	1.21	0.51	0.226
Q11	155	23.9	36.1	40.0	2.16	0.79	94	25.5	42.6	31.9	2.06	0.76	0.422
Q12	150	30.0	42.0	28.0	1.98	0.76	94	26.6	44.7	28.7	2.02	0.75	0.843
Q13	143	21.7	43.4	35.0	2.13	0.74	90	24.4	36.7	38.9	2.14	0.79	0.599
Q14	161	18.0	18.0	64.0	2.46	0.78	98	14.3	17.3	68.4	2.54	0.73	0.703
Q15	135	10.4	10.4	73.9	2.69	0.65	88	6.8	17.0	76.1	2.69	0.59	0.267
Q16	174	63.2	22.4	14.4	1.51	0.74	99	61.6	23.2	15.2	1.54	0.75	0.965
Q17	161	46.6	26.1	27.3	1.81	0.84	96	37.5	32.3	30.2	1.93	0.82	0.343

Appendix 10: Full data for comparison of levels of agreement between English universities and institutions from the other parts of the UK

Question	England n	England Disagree %	England Neither agree nor disagree %	England Agree %	England mean response	England standard deviation	Other UK n	Other UK Disagree %	Other UK Neither agree nor disagree %	Other UK Agree %	Other UK mean response	Other UK standard deviation	Sig
Q1	199	1.5	2.5	96.0	2.94	0.29	86	1.2	1.2	97.7	2.97	0.24	0.754
Q2	198	1.0	4.5	94.4	2.93	0.29	86	2.3	3.5	94.2	2.92	0.35	0.639
Q3	170	50.6	34.7	14.7	1.64	0.73	79	43.0	35.4	21.5	1.78	0.78	0.345
Q4	176	4.0	21.0	75.0	2.71	0.54	75	2.7	34.7	62.7	2.60	0.55	0.072
Q5	179	18.4	40.8	40.8	2.22	0.74	82	15.9	36.6	47.6	2.32	0.73	0.587
Q6	179	17.9	36.9	45.3	2.27	0.75	79	16.5	25.3	58.2	2.42	0.76	0.127
Q7	181	47.0	29.3	23.8	1.77	0.81	78	38.5	32.1	29.5	1.91	0.82	0.421
Q8	180	66.1	25.0	8.9	1.43	0.65	78	59.0	32.1	9.0	1.50	0.66	0.489
Q9	178	36.0	33.1	30.9	1.95	0.82	76	31.6	30.3	38.2	2.07	0.84	0.527
Q10	179	87.2	10.6	2.2	1.15	0.42	81	85.2	12.3	2.5	1.17	0.44	0.910
Q11	174	23.0	38.5	38.5	2.16	0.77	75	28.0	38.7	33.3	2.05	0.79	0.632
Q12	168	31.5	43.5	25.0	1.93	0.75	76	22.4	42.1	35.5	2.13	0.75	0.165
Q13	162	27.2	43.2	29.6	2.02	0.76	71	12.7	35.2	52.1	2.39	0.71	0.002
Q14	180	16.7	19.4	63.9	2.47	0.77	79	16.5	13.9	69.6	2.53	0.77	0.544
Q15	153	10.5	11.8	77.8	2.67	0.66	70	5.7	15.7	78.6	2.73	0.56	0.410
Q16	191	60.2	22.5	17.3	1.57	0.77	82	68.3	23.2	8.5	1.40	0.65	0.165
Q17	179	44.7	30.2	25.1	1.80	0.81	78	39.7	24.4	35.9	1.96	0.87	0.205