Coherence between text comments and the quantitative ratings in the UK’s National Student Survey.

A. Mark Langan¹, Nick Scott², Shobana Partington³ and Agnieszka Oczujda².

¹Faculty of Science and Engineering, Chester Street, Manchester Metropolitan University, M1 5GD, UK;
²Manchester Metropolitan University Business School, Manchester Metropolitan University
³Hollings Faculty, Righton Building, Cavendish Street, Manchester Metropolitan University

¹Author to whom correspondence should be addressed; m.langan@mmu.ac.uk, tel.: +44 (0)161 247583; fax: +44 (0)161 2476318.

Abstract

Institutions are understandably interested in the profile of their own reputations based upon publicly available data about student experiences. The UK’s National Student Survey (NSS) metrics are integrated into several ‘Good University’ calculations, whereas teaching teams most often use the survey’s text comments to change practices, rather than the metrics directly. There is little information about how messages from the national survey’s text comments relate to the accompanying numerical ratings, partly because text comments are confidential to the institution and unavailable for wide-scale research. We categorised institutional NSS text comments into themes that mirrored those of the original questionnaire. Comparisons were made between frequencies of thematic comments and the national ratings of satisfaction for several subject areas. For the first time we demonstrate broad agreement between comments about measures of teaching staff and course organisation with the performance of the subject areas (compared to metrics of their national counterparts). These findings are consistent with previous quantitative models predicting the most important factors that most influence overall satisfaction ratings. We intend this study to be a catalyst for other institutions to explore their non-publicly available, textual returns in a similar way. The outcomes of this type of work are pertinent to all countries that use large-scale surveys. However, institutions will need to release findings to a public audience if we are to gain a national/international perspective on this key linkage between publicly available metrics and the associated text comments.
Keywords: Student surveys, satisfaction, educational ratings
Introduction

For the past two decades, there has been considerable interest worldwide from Higher Education Institutions (HEIs) in the development and use of mass student surveys (Gibbs 2010; Buckley 2012). The UK’s National Student Survey (NSS) was designed to inform university applicants about their study choices and to provide a measure of institutional accountability at a national scale (Surridge 2009). However, it is clearly of benefit to institutions to perform well in the well-publicised survey, as HEIs compete for student applications and institutional prestige (Gibbs 2012). It is arguable that the impact of the national survey has increased beyond its intention and capabilities (see Langan et al. 2013) as its outcomes are being acted upon directly by HEIs in an attempt to improve their metrics as the priority, rather than a primary focus on the educational gains of its learners (Gibbs 2010; 2012). Apart from improvement of the learning experience quality itself, it is in the interest of institutions to enhance (directly) the survey metrics per se due to their public usage (Gibbs 2010). However, there are significant difficulties with the processes of contextualising the information generated by these large-scale survey instruments, from both local and national perspectives (Fielding et al. 2010, Langan et al. 2013).

The NSS (www.thestudentsurvey.com/content/nss2012_questionnaire_english.pdf) is carried out principally in the final year of undergraduate study and comprises 21 standard items (Q1-Q21; but note that an additional question about Students’ Unions is now included). These are designed to rate student perceptions of their experiences of the whole course. The survey targets several thematic areas; Teaching, Assessment & Feedback (often considered separately; e.g. Marsh & Cheng 2008), Academic Support, Organisation & Management, Learning Resources and Personal Development. There is also a global measure of satisfaction with the course (Q22) and opportunity to add in other institution-specific questions. These are followed by the request for free text comments to capture the respondent’s views of best practices and areas for improvement.

The questionnaire requires respondents to express strengths of agreement with the positive statements on a scale of 1 (strongly disagree) to 5 (strongly agree) with the central value of 3 indicating neutrality. Such an approach can be ambiguous and even misleading in terms of insights into the student experience (Blair et al. 2012). The NSS outcomes can highlight areas for targeted actions, but the metrics alone do not provide clarity around the nature of dissatisfaction (or non-satisfaction), which makes decision-making around learning design difficult (see Kovacs et al. 2010). Often, local decision-making is informed by the rich text comments provided by the survey, in conjunction with other quality enhancement procedures, to ultimately drive educational quality enhancement (Buckley 2012).

Despite its huge impact, there is little information about the congruence between the national survey’s quantitative metrics and the associated qualitative comments. Staff at a local level have raised concerns that there is a push to ‘enhance metrics’ by responding to the textual comments. However, as far as the authors’ are aware, there is currently is no evidence base in the literature that the comments and metrics in the UK’s survey are substantially related. The lack of evidence of this important link is partly because text comments are confidential to the institution and are not publicly
available for wider comparisons. With availability of both types of information at an institutional level, there is good reason to explore any associations (or disassociations) between the qualitative and quantitative survey outputs and to share the findings with the wider academic community. The current study is proposed as a useful, exploratory study highlighting broad patterns only, and is not intended to contribute to the debate about concepts such as triangulation, abductive logic or the epistemologies that are debated in detail elsewhere (e.g. Kelle 2001). The central purpose is to identify links between the survey’s text comments and the associated quantitative ratings, whilst retaining sufficient anonymity at an institutional level of the subject areas under scrutiny. We intend for this study to prompt discussion about the value of the survey which reveals its metrics publicly, but retains as confidential (at an institutional level) the valuable comments from respondents.

**Methods**

**Premise and approach**

Each year in the UK, NSS metrics are made available publicly via the Unistats website ([http://unistats.direct.gov.uk/](http://unistats.direct.gov.uk/)) but these are aggregated into broader subject areas (JACS3 level). Institutions participating in the NSS are provided with the anonymised text comments that are at a course level. This study used text comments from the 2011/12 National Student Survey returned to the home HEI. A thematic analysis was completed on the institutional NSS returns that assigned each student’s text comment to the thematic categories that are used to group the quantitative survey themes (such as ‘Teaching’ and ‘Learning Resources’; see Surridge 2009). Comments that related to other areas (such as university buildings or personal lives of respondents) were categorised as ‘Other’. The analysis generated frequencies of comments assigned to the categories.

The overwhelming majority of comments (>95%) were very straightforward to classify both in terms of the coding systems and whether they were positive or negative. A few comments were more ambiguous. We adopted a policy of reflecting the student's emphasis on the comment. As an example, if a student wrote “some lecturers are good” in the positive comment area then we would classify this as a positive comment. Arguably, this could also be interpreted as some were not but since the student emphasis in this case was positive so it was classified as such. Likewise, if a student commented "some lecturers are poor" in the negative comment box then we would classify that as a negative as the emphasis is on the poor aspects of some staff. When comments mixed positive and negative “the course resources were insufficient, but the lecturers were very good” both elements were recorded in the spirit of the emphasis. In this case negative for resources, but positive for teaching staff. The thematic analysis was carried out by two researchers and at the start of this process, these 'raters' were in contact regularly to standardise any areas of uncertainty.

Two themes of the NSS (‘Teaching’ and ‘Organisation and Management’) are known to be of particular importance in predicting (quantitative) overall satisfaction ratings at a national level (e.g. Fielding et al. 2010; Langan et al. 2013). The ‘Teaching’ dimension was split for the purposes of the study as a means to isolate comments about
members of staff from comments about the design of the curriculum. The dominance
of comments about individual staff in the survey returns and the large body of
research indicating how teachers are crucial in ratings of the teaching experience
(e.g. Marsh 2007) led to the focus being placed on comments about teaching staff.
Here we explore how well frequencies of positive and negative comments related to
the quantitative ratings of overall satisfaction (Q22) for courses (or groupings of
courses) compared to their national counterparts. For example, NSS metrics of
satisfaction for a degree in Law (note that this is not a subject included in the current
study) would be compared to national levels of satisfaction for Law courses.

Survey responses to three general subject areas were explored, but to retain
anonymity for institutional purposes, these are referred to as ‘Areas’ A, B and C. In
Areas A (with four networks of courses) and B (with three networks of courses), the
local organisation meant that data were aggregated in a way that did not allow
individual courses to be isolated and directly related to national Overall Satisfaction
(Q22) means. These two Areas are used in this study to show general patterns, i.e. the
subject groupings in networks that perform well, or not, compared to national
counterparts when the subject mix is considered together. It is noteworthy that there
were generally consistent performances within networks to allow this generalisation
to be made. For Area C, the specific courses were identifiable and thus had an exact
national mean to be compared against. This Area was used to explore the reliability of
indicator ratios (i.e. by multiplying the ratios of good/bad comments in ‘Teaching’ and
‘Organisation and Management’ themes) as correlates of national performance using
national averages of the metrics for the subjects.

Calculation of the ratios and correlations

The general premise is that the ratio of ‘positive versus negative’ comments about
staff would be higher in areas that perform well when compared to national
counterparts. It was always the case that more negative comments were provided for
the theme of Organisation & Management (O&M). The final ratio simply multiplied the
ratios of the two themes, giving them equal weighting. This meant higher ratios
express more positive comments about staff and less negative comments about
course organisation. There was also an issue with no comments being made about
some thematic areas in some cases. To account for zeros in the dataset, all
frequencies of comments had one added (+1 in all cases) to ensure a ratio could be
calculated. This allowed the calculation of ratios to be completed for all areas
surveyed (since some had zero comments), but please note that the modification did
not influence the ratings in the Likert scale (1-5) that the respondents completed in
the original survey. Thus, the final calculation of the final ratio was:

\[
\frac{\text{Frequency of positive comments} + 1 \ (\text{Teaching})}{\text{Frequency of negative comments} + 1 \ (\text{Teaching})} \times \frac{\text{Frequency of positive comments} + 1 \ (\text{Organisation & Management})}{\text{Frequency of negative comments} + 1 \ (\text{Organisation & Management})}
\]

Mean ratings of courses were compared to national means using publicly available
national data (provided on http://unistats.direct.gov.uk/) for the subject areas in the
institutional return of NSS data (which is not publicly available). Spearman’s rank
correlation coefficients were calculated to explore relationships between the
 calculated ratios and the deviation of the course from the national mean level of
 satisfaction. All analyses were carried out using SPSS v19.0 (IBM Inc, Chicago, Illinois,
 USA).

**Results**

Areas A and B were used in the first exploration in a general sense as the subjects
 were embedded into ‘networks’ of courses. For the purposes of this study (and to
 retain course anonymity), the term network is used to indicate a set of related courses
 within each Area. This was achieved by considering together all the subjects taught in
 the networks, and providing an indication (only) of their collective performance at a
 national level. In all cases, there was a general consistency in performance of the
 subjects in their groupings, making this collective approximation straightforward. For
 example, there were no cases of one particular subject performing well in a grouping
 where others in the same grouping performed poorly (data not shown). It was clear in
 Area A (Figure 1) that the numbers of positive comments about the teaching staff
 were much higher in subject areas that performed well and, conversely, there were
 fewer negative comments about course organisation.

These patterns were present in the other subject areas, reflecting their relative
 performances, but were not as pronounced in Area B (Figure 2).

These results are summarised in Table 1, which shows the absolute values of the
 ratios of the thematic comments and the broad indicators of the performance of these
 subjects groups against national indicators. <insert Table 1>

For Area C, the absolute values of national performance in the 2011/12 NSS for each
 of the JACS3 level subjects were available and these correlated positively with the
 calculated ratios ($r_s = 0.786$, $n = 7$, $p = 0.036$). Thus ratios of comments increased
 (reflecting generally more positive views of the respondents) in accordance with the
 subject performance nationally, reflecting higher frequencies of positive comments
 about teaching staff and fewer negative comments about course organisation for
 courses that were ‘performing well’ (see Figure 3). When carrying out correlations for
 the ratio of comments for each of the two items in the final ratio separately, both
 were positive but were not significant (for comments about staff; $r_s = 0.464$, $n = 7$, $p$
 = 0.294: and comments about course organisation; $r_s = 0.500$, $n = 7$, $p = 0.253$).

<insert Figure 3>

**Discussion**
Capturing the student voice and then responding effectively, has obvious value in the
evaluation and development of the success of courses. In a broad sense, this study
has shown coherence between the metrics and text comments for this independently
administered, national student survey. The patterns appear consistent across three
disparate subject areas. There is a difficulty with reporting findings like this due to
sensitivity of institutions about NSS data, primarily resulting from reporting of the
data in public and local domains. Different subject areas need to be considered
separately (Fielding et al. 2010) as they consistently receive different ratings of
positive to negative comments and this may reflect many things, such as their
cultural differences (see Gibbs 2012). We believe that ‘what is written in the survey
responses’ and ‘what scores are provided’ do broadly relate. This is based on
evidence from the combination of: (i) the broad patterns of higher ratios in subject
areas that ‘performed well’ (i.e. above the national average) within the two Areas of
networks of courses; and, (ii) the final curvilinear relationship when the ratios were
explored at course level. The use of the two themes (i.e. ratings of ‘Teaching’ and
‘Organisation and Management’ only) was triggered by their dominance in the text
comments, particularly in terms of comments about teaching, and their prevalence in
previous quantitative models (e.g. Langan et al. 2013). We anticipated that courses
that receive many positive comments about staff, and few comments about poor
course organisation would score better in their metrics (for exploration of teacher
effects on learner ratings of teaching see Marsh 2007). Primarily this was due to these
particular themes of the survey being important in the quantitative models that
predicted Overall Satisfaction (Q22; Langan et al. 2013). It is important to note that
these are comparisons with national counterparts in general subject areas, rather
than comparison against individual subjects. This was mostly due to the problems
with comparing metrics from different subject areas (for detail about subject
differences in metrics see Fielding et al. 2010; and Marsh & Cheng 2008; for an
exploration of interpretations of the survey items see Blair et al. 2012). It is not
straightforward to elucidate how the survey dimensions inter-relate. The ‘teachers’
could potentially have a strong influence on the respondents’ views of other aspects
of the survey (such as learning resources or assessment) and there is little
understanding of how the national survey’s responses (metrics and comments) reveal
true educational gains (rather than what learners like; see Gibbs 2010). The need for
in-depth explorations of these areas is apparent if decision-makers are to maximise
the usefulness of the survey.

The lack of significance detected for correlations that explored separately the ratios of
comments about staff and course organisation reinforces the need for multiple factors
to be considered simultaneously in the consideration of the complex area of ratings of
student satisfaction. It is notable that these analyses were limited in sample sizes due
to the numbers of areas that can be reported on in the survey (n = 7 in this case).
Also, the addition of 1 to all frequencies to allow ratios to be calculated introduces
some bias to the correlations since it influences low sample sizes more greatly than
higher those with higher frequencies (although the use of ranked correlations limits
this bias and generally the patterns are clearly visible in all Areas). It also underlines
the value of the initial observations of the data (i.e. that both ‘Teaching’ and
‘Organisation and Management’ are important as indicators when scrutinising the
frequencies of text comments returned shown in Figures 1 and 2) which is how they
were interpreted in a broad sense locally and prompted this study in the first instance. There is potential to develop this work to the other themes of the questionnaire, but this would need greater numbers of comments about other themes of the survey (since comments about teaching, staff etc. dominated the return). It would also require sensitive data (i.e. the text comments) to be released from institutions to evaluate how consistent these patterns are for other institutions and courses at a national level.

The potential to internally survey students in earlier years (and in the term before the NSS is released to final year students) provides a significant opportunity to identify ‘early warning indicators’ using the ratios described in the current study. Many institutions have their own surveys in place and many include statements derived from the NSS. Subject areas that suddenly change in terms of comments about staff or course organisation would be useful to identify and be used to generate dialogue between staff and students to resolve issues earlier in the student lifecycle, and to enhance student perceptions of their experiences in tertiary education.

We acknowledge that this study provides a basic interpretation of the survey returns, designed only to highlight broad patterns in the text responses and the performance of subjects. Although often used, counts of the number of times that text comments relate to qualitative codes provide only a simple approach to gauge major patterns in questionnaire returns (Driscoll et al. 2007). Despite this ‘data-reduction synthesis’ losing information about the subtle differences between the respondents’ comments, such analyses can detect general themes in responses and their popularity/value have led to qualitative data analysis software programs to carry out the thematic analysis. Such automation is prone to bias due to respondents who re-emphasise particular concepts (see Onwuegbuzie & Teddlie 2003). With sufficient research, such systems could be used for a rapid evaluation of areas that may be at risk of low levels of student satisfaction.

There is a need to build on the current findings to provide greater context to these conclusions. This should include debate about why certain aspects of the student experience have more influence on the ratings provided and which critical factors prompt certain ratings or comments to be recorded on the survey. This could be achieved by in-depth discussions with students surrounding the drivers of their responses, both in terms of text comments and metrics. This would require careful experimental design to protect students through anonymisation procedures, for example with researchers unconnected to their institution. In addition, further evaluations could be made about how students interpret the wording of questions in this survey (Blair et al. 2012) and how these interpretations differ from those of the academic staff who use the survey outcomes to enhance the quality of courses. A disadvantage commonly voiced by qualitative researchers is a loss of depth (and flexibility) that occurs when qualitative data are quantified (Driscoll 2007). Whereas the qualitative codes can provide insights into many interrelated conceptual themes (Bazeley 2004), the responses were reduced into a more ‘one-dimensional’ quantitative derivation by categorising them as simply positive or negative. Further work is required to clarify the underlying complexity behind apparently straightforward responses from simple numerical ratings (of 1-5) and the accompanying written comments that provide a much more detailed overview of the
experiences of the students. However, the findings are a first step to highlight areas
to target surrounding the student voice and this supports the quantitative metrics that
the respondents provided (Langan et al. 2013).

There is a large body of evidence that the NSS does not provide direct measures of
educational gains (Gibbs 2010) and also it is noted that the survey is not intended to
do this (Surridge 2009). In a slightly different area of study, Coates (2009) highlighted
the difficulty of directly measuring the outcomes of training courses at a national level
given the scale and diversity of training operations and contexts. This mirrors the
difficulty with using national level surveys such as the NSS to capture student
satisfaction at large scale, as this leads to output metrics being accepted without
sufficient context (see Langan et al. 2013). The free comments provide greater depth
and can inform those involved with the design of learning systems of areas to
improve, but should be seen as the start of this process rather than the only evidence
base (Buckley 2012).

The current study suggests that the ‘student voice’ captured in a national survey, in
the form of text comments about the staff teaching on the courses and the
organisation, relate in at least a broad sense to the questionnaire scores of overall
satisfaction. There is potential for the use of these ratios as early warning systems if
internal surveys are used to capture text comments during the full student lifecycle.
The techniques described in the current study could be applied to earlier surveys to
indicate areas that receive ‘poorer’ ratios of comments about teaching staff or course
organisation. The ratios from the national survey described here could also be used to
stimulate dialogue between students and staff about their learning experiences. This
is not to preclude discussion about other elements of the learning experience, but the
ratios could be used to reassure staff that links exist between these two forms of
survey data. The wording of survey items may also influence the nature of comments
that are volunteered by respondents, and it seems timely for other measures, such as
engagement (Trowler & Trowler 2010) and belonging (Blair et al. 2012) to be used
together with ratings of the experience to gain a better understanding of the complex
nature of learners’ experiences of tertiary education.

Decades of empirical research (Pascarella and Terenzini 2005; Kuh et al. 2008) have
affirmed that the active engagement of learners in effective training practices plays a
critical role in developing high-quality outcomes. Lessons from the current study could
be applied to qualitative comments from these alternative questionnaire designs such
that, in addition to new metrics, the respondents should have reflected upon their
senses of ‘belonging’ and/or ‘engagement’ before writing text comments. Future work
could then begin to link ratings of the student experience with (self-assigned)
measures of active involvement and feelings of belonging, using the current study as
one method for such an exploration. Provision of such information in surveys would
allow analysis of the associated comments after respondents have been asked to
consider wider aspects of their experiences and could be followed by in-depth
interviews with students to gain greater context as to why they responded as they did
and what can be changed to enhance their satisfaction and learning gains. This wider
view of the experience of university life, and the subsequent ways to process the
outcomes, provide the next challenge to researchers of the student experience.
Notes on contributors

Mark Langan is a Senior Fellow (Learning and Teaching) and a Principal Lecturer in the Faculty of Science and Engineering. He has interests in student assessment and the use of measures of the student experience and educational engagement. He has spoken and published widely about maximising the use of student surveys to enhance quality in higher education.

Nick Scott is a Senior Fellow (Learning and Teaching) and a Principal Lecturer in the Manchester Metropolitan University Business School. He manages and explores the student experience.

Shobana Partington is a Principal Lecturer in the Hollings Faculty who manages and explores student satisfaction.

Agnieska Oczujda is a Research Assistant in the Manchester Metropolitan University Business School exploring student satisfaction with interests in quality enhancement and assurance in Higher Education.

References


Blair, B., S. Orr, and M. Yorke. 2012. ‘Erm, that question... I think I probably would’ve just put something in the middle and sort of moved on to the next one, because I think it’s really unclear’: How art and design students understand and interpret the National Student Survey. Higher Education Academy, York, UK. http://www.heacademy.ac.uk/assets/documents/resources/publications/GLAD_FINALns.doc


Kovacs, S., L. Grant, and F. Hyland. 2010. Research papers in Education. A study of the use of the National Student Survey to enhance the student experience in Education Departments. ESCcalate- Subject Centre for Education. Higher Education Academy, York, UK.


