Assessment & Evaluation in Higher Education
Vol. 30, No. 4, August 2005, pp. 387–415

ISSN 0260-2938 (print)/ISSN 1469-297X (online)/05/040387–29

DOI: 10.1080/02602930500099193

students? Why are response rates important? And how seriously is student feedback taken?

Students' evaluations of teaching

In North America, the practice of obtaining student feedback on individual teachers and course units is widespread. Marsh and Dunkin (1992) identified four purposes for collecting students' evaluations of teaching (SETs):

- Diagnostic feedback to teachers about the effectiveness of their teaching.
- A measure of teaching effectiveness to be used in administrative decision making.
-

The test-retest reliability of students' evaluations is high, even when there is an extended period between the two evaluations. The interrater reliability of the average ratings given by groups of students is also high, provided that the average is based on 10 or more students. There is a high correlation between the ratings produced by students taking *different* course units taught by the *same* teacher, but little or no relationship between the ratings given by students taking the *same* course unit taught by *different* teachers. This suggests that students' evaluations are a function of the person teaching the course unit rather than the particular unit being taught.

Evaluations of the same teachers given by successive cohorts of students are highlyw(students t
s 1 0 TD 0 Tw (s7.641ent)Tj /F2 1 ocula 0 TD ( )Tj 01einour TD -0.0446 Tw9same

s in ined observthathnee Murray,y o83).TD -0.0195 Tw (Evaluations358t rell nrinciple, u.01n

unit where different groups of students are taught by different instructors but are subject to the same form of assessment. In these circumstances, there is a clear relationship between SETs and academic attainment, even when the grades are assigned by an independent evaluator, although some aspects of teaching are more important in predicting attainment than others (Cohen, 1981; Marsh, 1987).

The relationship between SETs and academic attainment is stronger when students know their final grades, though there is still a moderate correlation if they provide their ratings before their final grades are known (Cohen, 1981). Greenwald and Gilmore (1997a, b) noted that in the latter case the students can acquire expectations about their final grades from the results of midterm tests. They found a *positive* relationship between students' expected grades and their overall ratings of their teaching but a *negative* relationship between students' expected grades and their estimated workload. They argued that students reduced their work investment in order to achieve their original aspirations when faced with lenient assessment on their midterm tests.

The latter research raises the possibility that SETs might be biased by the effects of extraneous background factors, a possibility that is often used to foster scepticism about the value of SETs in the evaluation of teaching in higher education (Husbands & Fosh, 1993). Marsh (1987) found that four variables were potentially important in predicting SETs: the students' prior interest in the subject matter; their expected grades; their perceived workload; and their reasons for taking the course unit in question. Nevertheless, the effects of these variables upon students' ratings were relatively weak and did not necessarily constitute a bias. For instance, course units that were perceived to have a higher workload received more positive ratings, and the effect of prior interest was mainly on what students said they had learned from the course unit rather than their evaluation of the teaching *per se* (see Marsh, 1983).

Marsh (1987) acknowledged in particular that more positive SETs might arise from the students' satisfaction at receiving higher grades (the *grading satisfaction hypothesis*) or else from other uncontrolled characteristics of the student population. The fact that the relationship between SETs and academic attainment is stronger when the students know their final grades is consistent with the grading satisfaction hypothesis. However, Marsh pointed out that, if students are taught in different groups on the same course unit, they may know how their attainment compares with that of the other students in their group, but they have no basis for knowing how their attainment compares with that of the students in other groups. Yet the correlation between SETs and academic attainment arises even when it is calculated from the tee factGuinearelatSper Tj983).

translated) for different educational settings and in some of these studies a different response scale was used. Even so, in each case both the reliability and the validity of the SEEQ were confirmed.

In a trial carried out by the Curtin University of Technology Teaching Learning Group (1997), the SEEQ was found to be far more acceptable to teachers than the existing in-house instrument. Coffey and Gibbs (2001) arranged for a shortened version of the SEEQ containing 24 items from six scales) to be administered to students at nine universities in the UK. The results confirmed the intended factor structure of this inventory and also showed a high level of internal consistency. Because cross-cultural research tended to confirm the factor structure of the SEEQ, Marsh and Roche (1994) argued that it was especially appropriate for the increasingly multicultural student population attending Australian universities.

In a further study, Coffey and Gibbs (in press) asked 399 new teachers from eight countries to complete a questionnaire about their approaches to teaching. They found that those teachers who adopted a student-focused or learning-centred approach to teaching received significantly higher ratings from their students on five of the six scales in the shortened SEEQ than did those teachers who adopted a teacher-focused or subject-centred approach to teaching. In the case of teachers who had completed the first semester of a training programme, Coffey and Gibbs (2000) found that their students gave them significantly higher ratings on four of the six scales in the shortened SEEQ at the end of the semester than they had done after four weeks. Nevertheless, this study suffered from a severe attrition of participants, and it is possible that the latter effect was simply an artefact resulting from sampling bias. Equally, the students may have given more positive ratings simply because they were more familiar with their teachers.

SETs are most commonly obtained when teaching is face-to-face and is controlled by a single lecturer or instructor. It has indeed been suggested that the routine use of questionnaires to obtain students' evaluations of their teachers promotes an uncritical acceptance of traditional conceptions of teaching based on the bare transmission of knowledge and the neglect of more sophisticated conceptions concerned with the promotion of critical thinking and self-expression (Kolitch & Dean, 1999). It should be possible to collect SETs in other teaching situations such as the supervision of research students, but there has been little or no research on the matter.

A different situation is that of distance education, where students are both physically and socially separated from their teachers, from their institutions, and often from other students too (Kahl & Cropley, 1986). To reduce what Moore (1980) called the 'transactional distance' with their students, most distance-learning institutions use various kinds of personal support, such as tutorials or self-help groups arranged on a local basis, induction courses or residential schools, and teleconferencing or computer conferencing. This support seems to be highly valued by the students in question (Hennessy *et al.*, 1999; Fung & Carr, 2000). Nevertheless, it means that 'teachers' have different roles in distance education: as authors of course materials and as tutors. Gibbs and Coffey (2001) suggested that collecting SETs in distance

education could help to clarify the expectations of both tutors and students about the nature of their relationship.

The intellectual rights and copyright in the SEEQ belong to Professor Herbert W. Marsh of the University of Western Sydney, Macarthur. It is presented on a double-sided form that allows for the inclusion of supplementary items and open-ended questions. If the SEEQ is administered in a class setting, respondents may be asked to record the course unit and the teacher being rated, but they themselves can remain anonymous. Marsh and Roche (1994) elaborated the SEEQ as the core of a self-

A similar approach has been adopted in in-house satisfaction surveys developed in the UK, but most have of these have not been adequately documented or evaluated. Harvey *et al.* (1997) described a general methodology for developing student satisfaction surveys based upon their use at the University of Central England. First, significant aspects of students' experience are identified from the use of focus groups. Second, these are incorporated into a questionnaire survey in which larger samples of students are asked to rate their satisfaction with each aspect and its importance to their learning experience. Finally, the responses from the survey are used to identify aspects of the student experience that are associated with high levels of importance but low levels of satisfaction. According to Harvey (2001), this methodology has been adopted at a number of institutions in the UK and in some other countries, too. Descriptive data from such surveys have been reported in institutional reports (see Harvey, 1995), but no formal evidence with regard to their reliability or validity has been published.

Students' perceptions of academic quality

From the perspective of an institution of higher education seeking to maintain and improve the quality of its teaching, it could be argued that the appropriate focus of assessment would be a programme of study rather than an individual course unit or the whole institution, and this has been the dominant focus in Australia and the UK.

In an investigation into determinants of approaches to studying in higher education, Ramsden and Entwistle (1981) developed the Course Perceptions Questionnaire (CPQ) to measure the experiences of British students in particular degree programmes and departments. In its final version, the CPQ contained 40 items in eight scales that reflected different aspects of effective teaching. It was used by Ramsden and Entwistle in a survey of 2208 students across 66 academic departments of engineering, physics, economics, psychology, history and English. A factor analysis of their scores on the eight scales suggested the existence of two underlying dimensions: one reflected the positive evaluation of teaching and programmes, and the other reflected the use of formal methods of teaching and the programmes' vocational relevance.

The CPQ was devised as a research instrument to identify and to compare the perceptions of students on different programmes, and Ramsden and Entwistle were able to use it to reveal the impact of contextual factors on students' approaches to learning. However, the primary factor that underlies its constituent scales is open to a natural interpretation as a measure of perceived teaching quality, and Gibbs *et al.* (1988, pp. 29–33) argued that the CPQ could be used for teaching evaluation and course review. Even so, the correlations obtained by Ramsden and Entwistle between students' perceptions and their approaches to studying were relatively weak. Similar results were found by other researchers (Parsons, 1988) and this led to doubts being

on particular academic programmes. In the light of preliminary evidence, a national trial of the CEQ was commissioned by a group set up by the Australian Commonwealth Department of Employment, Education and Training to examine performance indicators in higher education (Linke, 1991). In this national trial, usable responses to the CEQ were obtained from 3372 final-year undergraduate students at 13 Australian universities and colleges of advanced education (see also Ramsden, 1991b).

The instrument used in this trial consisted of 30 items in five scales which had been identified in previous research as reflecting different dimensions of effective instruction: good teaching (8 items); clear goals and standards (5 items); appropriate workload (5 items); appropriate assessment (6 items); and emphasis on independence (6 items). The defining -1.237i j,in five scal(accordaining i j,us ult1.237i j, a national tr) a weshows on T usa 11). addi nat,7i r(s237i j,30 items by th; appropriaAe assman237rotableTw(resstudenEQ weve instt usng ce indie by ir leverial agr(artmenrentsagr(-een)T artmech a descripucati237i jir ic prograi237te sy. Halfi237i j(6
it
responsbiaponseithighng

GCCA)

overall response rates of around 60% (Ainley & Long, 1995; Johnson *et al.*, 1996; Johnson, 1997, 1998, 1999; Long & Hillman, 2000). However, in the GCCA surveys, the original version of the CEQ has been modified in certain respects:

- In response to concerns about the employability of graduates, a Generic Skills scale was added to 'investigate the extent to which higher education contributes to the enhancement of skills relevant to employment' (Ainley & Long, 1994, p. xii). This contains six new items that are concerned with problem solving, analytic skills, teamwork, communication and work planning. Of course, similar concerns about the process skills of graduates have been expressed in the UK (Committee of Vice-Chancellors and Principals, 1998). The items in the Generic Skills scale are somewhat different from those in the rest of the CEQ, insofar as they ask respondents to evaluate the skills that they have gained from their programmes rather than the quality of the programmes themselves. Other researchers have devised more extensive instruments for measuring graduates' perceptions of their personal development during their programmes of study (see Purcell & Pitcher, 1998; Cheng, 2001).
- To compensate for this and reduce the lengcom1iE2 insofhlie Tf7 0 0ltls furtr, 19e

research carried out in individual universities in Australia (Trigwell & Prosser, 1991) and Britain (Richardson, 1994). Evidence concerning the psychometric properties of the 23-item version of the CEQ has been obtained in the GCCA surveys and in the study by Wilson *et al.* (1997); the latter also provided evidence concerning the psychometric properties of the 36-item version of the CEQ.

The internal consistency of the scales as measured by Cronbach's (1951) coefficient alpha is generally satisfactory, although there is no evidence on their test–retest reliability. The composition of the scales according to the results of factor analyses conducted on the responses to individual items is broadly satisfactory. In the 23-item version, all of the items tend to load on distinct factors reflecting their assigned scales (see Byrne & Flood, 2003). The application of Rasch's (1960) measurement analysis confirms the multidimensional structure of the CEQ (Waugh, 1998; Ainley, 1999). In the 30-item and the 36-item versions, most items load on factors reflecting their assigned scales, but there is a consistent tendency for a few items on the Good Teaching scale and the Emphasis on Independence scale to load on other factors.

Two recent studies have identified a possible problem with the Good Teaching scale. Broomfield and Bligh (1998) obtained responses to the version of the CEQ devised by Ainley and Long (1994) from 180 medical students. A factor analysis confirmed the scale structure of this instrument, except that the Good Teaching scale was reflected in two separate factors: one was defined by three items concerned with the classroom instruction; the other was defined by two items concerned with feedback given to the students on their work. Kreber (2003) obtained similar results when she asked 1,080 Canadian students to evaluate particular course units. Of course, the quality of instruction is likely to depend on the competence of individual teachers, but the quality of feedback on students' work is likely to depend more on institutional practices.

The construct validity of the CEQ according to factor analyses on respondents' scores on the constituent scales is also broadly satisfactory. The modal solution is a single factor on which all of the scales show significant loadings. The Appropriate Workload scale shows the lowest loadings on this factor, and there is debate over whether it should be taken to define a separate dimension (Ainley, 1999; Richardson, 1997). The criterion validity of the CEQ as an index of perceived quality can be tested by examining the correlations between respondents' scale scores and their responses to the additional item concerned with their overall satisfaction. Typically, all of the CEQ's scales show statistically significant correlations with ratings of satisfaction (see also Byrne & Flood, 2003), but the Appropriate Workload scale shows the weakest associations.

The discriminant validity of the CEQ is shown by the fact that the respondents' scores on the constituent scales vary across different academic disciplines and across different institutions of higher education offering programmes in the same discipline. In particular, students produce higher scores in departments that pursue student-centred of experiential curricula through such models as problem-based learning (see also Eley, 1992; Sadlo, 1997). Conversely, Ainley and Long (1995) used results from the 1994 GCCA survey to identify departments of psychology in which there was 'the

possible need for review of teaching and assessment practices' (p. 50). Long and Hillman (2000, pp. 25–29) found in particular that ratings on the Good Teaching scale as well as students' overall level of satisfaction varied inversely with the size of their institution.

As mentioned earlier, Ramsden and Entwistle (1981) were originally concerned to demonstrate a connection between students' perceptions of their programmes and the approaches to learning that they adopted on those programmes. The weak relationships that they and other researchers found cast doubt upon the concurrent validity of the CPQ. In contrast, investigations carried out at the Open University have shown an intimate relationship between the scores obtained on the CEQ by

undergraduate programmes in occupational therapy at institutions of higher education in six different countries. In the UK, the 30-item version of the CEQ has been used both for academic review (Richardson, 1994) and for course development (Gregory *et al.*, 1994, 1995). Wilson *et al.* (1997) advised that the CEQ was not intended to provide feedback with regard to individual subjects or teachers. Nevertheless, Prosser *et al.* (1994) adapted the CEQ to refer to particular topics (such as mechanics in a physics programme or photosynthesis in a biology programme), and a modified version of the CEQ concerned with students' perceptions of individual course units has been used to compare their experience of large and small classes (Gibbs & Lucas, 1996; Lucas *et al.*, 1997). The Curtin University of Technology Teaching Learning Group (1997) reworded the 23-item version of the CEQ to refer to the lecturer teaching a specific course unit, and they proposed that it might complement the SEEQ in the evaluation of individual lecturers.

Mazuro *et al.* (2000) adapted the original 30-item version of the CEQ so that it could be completed by teachers with regard to their own course units. They gave this instrument to five members of academic staff teaching a course unit in social psychology, and they also obtained responses to the CEQ from 95 students who were taking the course unit in question. Mazuro *et al.* found that the correlation coefficient between the mean responses given by the staff and the students across the individual items of the CEQ was +0.54, suggesting that overall there was a high level of consistency between the teachers' and the students' perceptions of the course unit. The students gave more favourable ratings than the teachers on two items concerned with their freedom to study different topics and develop their own interests, but they gave less favourable ratings than the teachers on four items concerned with the quality of feedback, the amount of workload and the teachers' understanding of difficulties they might be having with their work.

The intellectual rights and the copyright in the CEQ belong to Professor Paul Ramsden (now the Chief Executive of the UK Higher Education Academy), the Graduate Careers Council of Australia and the Australian Commonwealth Department of Education, Training and Youth Affairs. Like the SEEQ, it can be conveniently presented on a double-sided form and the responses automatically scanned. In the GCCA surveys, a descriptive summary of the average ratings given to eachheir own f r

In response, a separate instrument, the Postgraduate Research Experience Questionnaire was developed (Johnson, 1999, p. 11). Initial results with this instrument indicated that it had reasonable internal consistency and a consistent structure based on six dimensions: supervision, skill development, intellectual climate, infrastructure, thesis examination, and goals and expectations. This instrument is now employed across the Australian university system, and the findings are returned to institutions but are not published.

Nevertheless, further research demonstrated that the questionnaire did not discriminate among different universities or among different disciplines at the same university (Marsh *et al.*, 2002). As a result, there is considerable scepticism about whether it provides an adequate basis for benchmarking universities or disciplines within universities. One difficulty is the lack of a coherent research base on the experiences of postgraduate research students, and this has encouraged the use of totally ad hoc instruments to measure their perceptions of quality. Another difficulty is that evaluations of research training typically confound the overall quality of the research environment with the practice of individual supervisors. It is only very recently that researchers and institutions have recognized the need to distinguish institutional monitoring from enhancing supervisory practice (Chiang, 2002; Pearson *et al.*, 2002).

The GCCA surveys also embrace students who have studied by distance education, for whom items referring to 'lecturers' or 'teaching staff' might be inappropriate. As mentioned earlier, academic staff in distance-learning institutions have two rather different roles: as the authors of course materials and as course tutors. Richardson and Woodley (2001) adapted the CEQ for use in distance education by amending any references to 'lecturers' or to 'teaching staff' so that the relevant items referred either to teaching materials or to tutors, as appropriate. The amended version was then used in a postal survey of students with and without a hearing loss who were taking course units by distance learning with the Open University. A factor analysis of their responses confirmed the intended structure of the CEQ, except that the Good Teaching scale split into two scales concerned with good materials and good tutoring. Similar results were obtained by Lawless and Richardson (2002) and by Richardson and Price (2003), suggesting that this amended version of the CEQ is highly robust in this distinctive context.

The CEQ was intended to differentiate among students taking different programmes of study, but the GCCA surveys have also identified apparent differences related to demographic characteristics of the respondents, including gender, age, first language and ethnicity. However, the authors of the annual reports from the GCCA surveys have been at pains to point out that these effects could simply reflect the enrolment of different kinds of student on programmes in different disciplines with different teaching practices and different assessment requirements. In other words, observed variations in CEQ scores might arise from respondents taking different programmes rather than from inherent characteristics of the respondents themselves. Indeed, in research with Open University students taking particular course units (Richardson, 2005; Richardson & Price, 2003), demographic characteristics

such as gender and age did not show any significant relationship with students' perceptions of the academic quality of their courses.

One potential criticism of the CEQ is that it does not include any items relating to the pastoral, physical or social support of students in higher education. It is entirely possible to include additional items concerned with institutional facilities, such as computing and library resources. In fact, some institutions involved in the Australian graduate surveys have included extra items regarding administrative matters, student services and recreational facilities, but these additional items were not considered in the published analysis of results from the CEQ (Johnson *et al.*, 1996, p. 3). An initial analysis suggested that students' satisfaction with their facilities was a much weaker prediction of their overall satisfaction than the original scales in the CEQ (Wilson *et al.*, 1997). As Johnson *et al.* (1996, p. 5) noted, the CEQ does not claim to be comprehensive but seeks information about dimensions of teaching and learning that appear to be central to the majority of academic subjects taught in institutions of higher education.

Nevertheless, further developments were motivated by discussions in focus groups with stakeholders as well as analyses of students' responses to open-ended questions included in the CEQ. McInnis *et al.* (2001) devised six new scales, each containing five items, to measure the domains of student support, learning resources, course organization, learning community, graduate qualities and intellectual motivation. The Course Organization scale proved not to be satisfactory, but McInnis *et al.* suggested that the other five scales could be used by institutions in annual surveys of their graduates. This would yield an extended CEQ containing 50 items. McInnis *et al.* found that students' scores on the new scales were correlated with their scores on the five original scales of the 23-item CEQ, and they concluded that the inclusion of the new scales had not affected their responses to the original scales (p. x; see also Griffin *et al.*, 2003).

However, McInnis *et al.* did not examine the constituent structure of their extended instrument in any detail. They have kindly provided a table of correlation coefficients among the scores of 2316 students on the five original scales of the 23-item CEQ and their six new scales. A factor analysis of the students' scores on all 11 scales yields a single underlying dimension, but this is mainly dominated by the new scales at the expense of the original scales. This might suggest that the extended 50-item version of the CEQ is perceived by students as being mainly concerned with informal aspects of higher education (such as resources and support systems). Like the Generic Skills scale, the new scales were introduced for largely pragmatic reasons and are not grounded in research on the student experience. Hence, although the extended CEQ taps a broader range of students' opinions, it may be less appropriate for measuring their perceptions of the more formal aspects of the curriculum that are usually taken to define teaching quality.

As with students' evaluations of teaching, there is little evidence that the collection of student feedback using the CEQ in itself leads to any improvement in the perceived quality of programmes of study. Even so, the proportion of graduates who agreed that they were satisfied with their programmes of study in the GCCA surveys has

gradually increased from 60% in 1995 to 68% in 2001, while the proportion who disagreed decreased from 14% to 10% over the same period (Graduate satisfaction, 2001). By analogy with the limited amount of evidence on the value of SETs, students' scores on the CEQ might assist in the process of course development, especially if used in a systematic process involving consultation and counselling (Gregory *et al.*, 1994, 1995), and they might also be expected to improve following specific interventions aimed at improving the quality of teaching and learning across entire programmes of study.

Practical issues in obtaining student feedback

Why obtain student feedback?

As mentioned earlier, student feedback can provide diagnostic evidence for teachers and also a measure of teaching effectiveness for administrative decision-making. In the UK, it is becoming increasing accepted that individual teachers will refer to student feedback both to enhance the effectiveness of their teaching and to support applications for appointment, tenure or promotion. Student feedback also constitutes information for prospective students and other stakeholders in the selection of programmes or course units, and it provides relevant evidence for research into the processes of teaching and learning. Clearly, both students' evaluations of teaching and their perceptions of academic quality have been investigated with each of these aims in mind. The research literature suggests that student feedback constitutes a major source of evidence for assessing teaching quality; that it can be used to inform attempts to improve teaching quality (but simply collecting such feedback is unlikely to lead to such improvements); and that student feedback can be communicated in a way that is informative to future students.

Why use formal instruments?

Student feedback can be obtained in many ways other than through the administration of formal questionnaires. These include casual comments made inside or outside the classroom, meetings of staff–student committees and student representation on institutional bodies, and good practice would encourage the use of all these means to maintain and enhance the quality of teaching and learning in higher education. However, surveys using formal instruments have two advantages: they provide an opportunity to obtain feedback from the entire population of students; and they document the experiences of the student population in a more or less systematic way.

One could obtain student feedback using open-ended questionnaires. These might be particularly appropriate on programmes in education, the humanities and the social sciences, where students are often encouraged to be sceptical about the value of quantitative methods for understanding human experience. Nevertheless, the burden of analyzing open-ended responses and other qualitative data is immense, even with only a relatively modest sample. The process of data analysis becomes quite

intractable with larger samples unless there are a limited number of response alternatives to each question that can be encoded in a straightforward way. The use of quantitative inventories to obtain student feedback has therefore been dictated by organizational constraints, particularly given the increasing size of classes in higher education. The content of such instruments could, of course, be based on results from qualitative research, as was the case for the CEQ, or from focus groups, as in Harvey *et al*'s (1997) student satisfaction methodology.

In addition, informal feedback is mainly available when teachers and learners are involved in face-to-face situations. In distance education, as mentioned earlier, students are both physically and socially separated from their teachers and their institutions, and this severely constrains the opportunities for obtaining student feedback. In this situation, the use of formal inventories has been dictated by geographical factors as much as by organizational ones (Morgan, 1984). It can be argued that it is not appropriate to compare the reports of students at institutions (such as the Open University) which are wholly committed to distance education with the reports of students at institutions which are wholly committed to face-to-face education. Nevertheless, it would be both appropriate and of theoretical interest to compare the reports of distance-learning and campus-based students taking the same programmes at the large number of institutions that offer both modes of course delivery, bearing in mind, of course, the obvious differences in the educational context and the student population (see Richardson, 2000).

What should be the subject of the feedback?

Student feedback can be obtained on teachers, course units, programmes of study, departments and institutions. At one extreme, one could envisage a teacher seeking feedback on a particular lecture; at the other extreme, one might envisage obtaining feedback on a national system of higher education, especially with regard to controversial developments such as the introduction of top-up fees. Nevertheless, it is clearly sensible to seek feedback at a level that is appropriate to one's basic goals. If the aim is to assess or improve the quality of particular teachers, they should be the subject of feedback. If the aim is to assess or improve the quality of particular programmes, then the latter should be the subject of feedback. Logically, there is no reason to think that obtaining feedback at one level would be effective in monitoring or improving quality at some other level (nor any research evidence to support this idea, either). Indeed, identifying problems at the programme or institutional level might have a negative impact on the quality of teaching by demotivating the staff who are actually responsible for delivering the programmes.

What kind of feedback should be collected?

Most of the research evidence has been concerned with students' perceptions of the quality of the teaching that they receive or their more global perceptions of the academic quality of their programmes. Much less evidence has been concerned with

students' level of satisfaction with the teaching that they receive or with their programmes in general. Consumer theory maintains that the difference between consumers' expectations and perceptions determines their level of satisfaction with the quality of provision of a service. This assumption is embodied in American instruments such as the Noel-Levitz Student Satisfaction Inventory and also in Harvey *et al*'s (1997) student satisfaction methodology. (Indeed, one could also modify the

respondents themselves and that earlier feedback would be of more immediate value. Indeed, Greenwald and Gilmore (1997a, b) found that students' perceptions in the middle of a course unit influenced their subsequent studying and final grades.

Others have suggested that the benefits or otherwise of having completed a programme of study are not immediately apparent to the new graduates, and hence feedback should be sought some time after graduation. Indeed, from a purely practical point of view, it would be both convenient and economical to obtain feedback from recent graduates at the same time as enquiring about their entry into employment or postgraduate education. In the United Kingdom the latter enquiries are made as part of the First Destination Survey (FDS). Concern has been expressed that this might reduce the response rate to the FDS and thus impair the quality of the information that is available about graduate employment (Information on quality, 2002, p. 15). However, the converse is also possible: incorporating the FDS might reduce the response rate to a survey of graduates' perceptions of the quality of their

has been successfully used with other problem-based programmes in higher education (Sadlo & Richardson, 2003; Trigwell & Prosser, 1991).

In the GCCA surveys, the CEQ seems to be appropriate for assessing the experience of students on both undergraduate and postgraduate programmes. (Students taking joint degree programmes are asked to provide responses for each of their disciplines separately.) However, it does not seem to be useful for assessing the experiences of students working for postgraduate research degrees, and no suitable alternative has yet been devised. It may prove necessary to evaluate the quality of postgraduate research training using a different methodology from the CEQ. In distance education, it has proved necessary to amend the wording of many of the items in the CEQ, and the constituent structure of the resulting questionnaire reflects the different roles of staff as the authors of course materials and as tutors. The wording and the structure of any instrument adopted for use in a national survey of graduates would have to accommodate the different practices in campus-based and distance education. More generally, it would have to be able to accommodate variations in practice in higher education that might arise in the future.

Why are response rates important?

Some might argue that the purpose of feedback surveys was simply to provide students with an opportunity to comment on their educational experience. On this argument, students who do not respond do not cause any difficulty because they have chosen not to contribute to this exercise. Nevertheless, most researchers assume that the purpose of feedback surveys is to investigate the experience of all the students in question, and in this case those who do not respond constitute a serious difficulty insofar as any conclusions have to be based on data contributed by a sample.

Inferences based upon samples may be inaccurate for two reasons: sampling error and sampling bias. Sampling error arises because, even if a sample is chosen entirely at random, properties of the sample will differ by chance from those of the population from which the sample has been drawn. In surveys, questionnaire responses generated by a sample will differ from those that would be generated by the entire population. The magnitude of the sampling error is reduced if the size of the sample is increased, and so efforts should be made to maximize the response rate. Sampling bias arises when a sample is not chosen at random from the relevant population. As a result, the properties of the sample may be misleading estimates of the corresponding properties of the population as a whole. In surveys, sampling bias arises if relevant characteristics of the people who respond are systematically different from those of the people who do not respond, in which case the results may be at variance with those that would have been found if responses had been obtained from the entire population.

Research has shown that people who respond to surveys are different from those who do not respond in terms of demographic characteristics such as age and social class (Goyder, 1987, chapter 5). In market research, a common strategy is to weight the responses of particular groups of respondents to compensate for these

with disabilities, and this is arguably an obligation under legislation such as the Americans with Disabilities Act in the US or the Special Educational Needs and Disability Act in the UK.

To achieve high response rates, it is clearly necessary to ensure the cooperation and motivation of the relevant population of students. Those who have satisfactorily completed a course unit or an entire programme may be disposed to complete feedback questionnaires, but this may not be the case for students who have failed and particularly for those who have withdrawn from their studies for academic reasons. At the Open University, students who drop out of course units are automatically sent a questionnaire to investigate the reasons for their withdrawal. This provides useful information, but the response rates are typically of the order of 25%. One could therefore not be confident that the data were representative of students who withdraw from course units.

How seriously is student feedback taken?

It is often assumed that the publication of student feedback will help students to make decisions about the choice of programmes and course units, that it will help teachers to enhance their own professional skills and that it will help institutions and funding bodies to manage their resources more effectively. None of these assumptions has been confirmed by empirical research, though it should be noted that most of the evidence relates to the use that is (or is not) made of SETs.

There have been consistent findings that students believe SETs to be accurate and important, although they constitute only one of the sources of information that students use when choosing among different course units (Babad, 2001). However, students may be sceptical as to whether attention is paid to the results either by the teachers being assessed or by senior staff responsible for appointments, appraisal or promotions, because they perceive that teachers and institutions attach more importance to research than to teaching. Indeed, unless students can see that the expression of their opinions leads to concrete changes in teaching practices, they may make little use of their own ratings (Spencer & Schmelkin, 2002). However, the development needs that students ascribe to their teachers may be driven by a didactic model of teaching and may differ from the teachers' own perceived needs (Ballantyne *et al.*, 2000).

From the teachers' perspective, the situation is a similar one. In the past, some resistance to the use of student ratings has been expressed based on the ideas that students are not competent to make such judgements or that student ratings are influenced by teachers' popularity rather than their effectiveness. Both sociability and

Even in institutions where the collection of student feedback is compulsory, teachers may make little attempt to make use of the information that it contains. Once again, this may be because institutions are perceived to attach more importance to research than to teaching, despite having formal policies that implicate teaching quality in decisions about staff appointments, appraisal and promotions (Kember *et al.*, 2002). There seems to be no published research evidence on the use that senior managers of institutions make or do not make of student feedback in such cases, but there are four main reasons for the apparent lack of attention to this kind of information.

The first reason is the lack of guidance to teachers, managers and administrators on how such information should be interpreted. In the absence of such guidance, there is little or no scope for any sensible discussion about the findings. Potential users of student feedback need to be helped to understand and contextualize the results (Neumann, 2000). The second reason is the lack of external incentives to make use of such information. In the absence of explicit rewards for good feedback or explicit penalties for poor feedback (or at least for not acting upon such feedback), it is rational for both teachers and students to infer that their institutions do not take the quality of teaching seriously and value other kinds of activities such as research (Kember *et al.*, 2002).

A third point is that the results need to be published to assure students that action is being taken, although care should also be taken that to ensure they are not misinterpreted or misrepresented. The Australian Vice-Chancellor's Committee (2001) issued a code of practice on the release of CEQ data, and this cautions against making simplistic comparisons among institutions (because of variations in the student populations at different institutions), aggregating the results from different disciplines to an institutional level (because of variations in the mix of disciplines at different institutions) and attaching undue importance to trivial differences in CEQ scores. In the UK, it would arguably be appropriate to report student feedback data for each institution in the 19 broad subject groupings used by the Higher Education Statistics Agency.

The final reason for the lack of attention to student feedback is the under-researched issue of the ownership of feedback data. Teachers may be less disposed to act on the findings of feedback, and students may be more disposed to be sceptical about the value of providing feedback to the extent that it appears to be divorced from the immediate context of teaching and learning. This is more likely to be the case if student feedback is collected, analyzed and published by their institution's central administration and even more so if it is collected, analyzed and published by an impersonal agency that is wholly external to their institution. The collection of feedback concerning programmes or institutions for quality assurance purposes certainly does not reduce the need to obtain feedback concerning teachers or course units for developmental purposes.

Conclusions

Surveys for obtaining student feedback are now well established both in North America and in Australia. They have also become increasingly common in the UK and are about

to become a fixture at a national level. Following the demise of subject-based review of the quality of teaching provision by the Quality Assurance Agency, a task group was set up in 2001 by the Higher Education Funding Council for England to identify the kinds of information that higher education institutions should make available to prospective students and other stakeholders. The group's report (commonly known as the 'Cooke report') proposed, amongst other things, that there should be a national survey of recent graduates to determine their opinions of the quality and standards of their experience of higher education (Information on quality, 2002, p. 15).

A project was set up by the Funding Council to advise on the design and administration of such a survey, which led in turn to the commissioning of a pilot study for a

may well be lower for students who have failed or who have withdrawn from their course units or programmes.

- Many students and teachers believe that student feedback is useful and informative, but many teachers and institutions do not take student feedback sufficiently seriously. The main issues are: the interpretation of feedback; institutional reward structures; the publication of feedback; and a sense of ownership of feedback on the part of both teachers and students.

Acknowledgements

This is a revised version of an article that was originally published in *Collecting and using student feedback on quality and standards of learning and teaching in HE*, available on the Internet at the hefce website under 'Publications/R&D reports'. It is reproduced here with permission from the Higher Education Funding Council for England. I am very grateful to John Brennan, Robin Brighton, Graham Gibbs, Herbert Marsh, Keith Trigwell, Ruth Williams and two anonymous reviewers for their various comments on earlier drafts of this article. I am also grateful to Hamish Coates for providing data from the study reported by McInnis *et al.* (2001).

Note on contributor

John T. E. Richardson is Professor of Student Learning and Assessment in the Institute of Educational Technology tional_Textx013hessor of Studsueducation2 d4t M

Australian Vice-Chancellor's Committee & Graduate Careers Council of Australia (2001) *Code of practice on the public disclosure of data from the Graduate Careers Council of Australia's graduate destination survey, Course Experience Questionnaire and postgraduate research experience questionnaire* (Canberra, Australian Vice-Chancellor's Committee). Available online at: www.avcc.edu.au (accessed 28 November 2002).

Babad, E. (2001) Students' course selection: differential considerations for first and last course, *Research in Higher Education*, 42, 469–492.

Babbie, E. R. (1973) *Survey research methods* (Belmont, CA, Wadsworth).

Ballantyne, R., Borthwick, J. & Packer, J. (2000) Beyond student evaluation of teaching: identifying and addressing academic staff development needs, *Assessment and Evaluation in Higher*

Gibbs, G. & Lucas, L. (1996) Using research to improve student learning in large classes, in: G. Gibbs (Ed.) *Improving student learning: using research to improve student learning* (Oxford, Oxford Centre for Staff Development).

Goyder, J. (1987) *The silent minority: nonrespondents on sample surveys* (Cambridge, Polity Press).

Graduate satisfaction (2001) *GradStats*, Number 6. Available online at: www.gradlink.edu.au/gradlink/gcca/GStats2001.pdf (accessed 27 November 2002).

Greenwald, A. G. & Gilmore, G. M. (1997a) Grading leniency is a removable contaminant of student ratings, *American Psychologist*, 52, 1209–1217.

Greenwald, A. G. & Gilmore, G. M. (1997b) No pain, no gain? The importance of measuring course workload in student ratings of instruction, *Journal of Educational Psychology*, 89, 743–751.

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

Gregory, R., Thorley, L. & Harland, G. (1994) Using a standard student experience questionnaire with engineering students: initial results, in: G. Gibbs (Ed.) *Improving student learning: theory and practice* (Oxford, Oxford Centre for Staff Development).

Griffin, P., Coates, H., McInnis, C. & James, R. (2003) The development of an extended Course Experience Questionnaire, *Quality in Higher Education*, 9, 259–266.

Harvey, L. (2001) *Student feedback: a report to the Higher Education Funding Council for England* (Birmingham, University of Central England, Centre for Research into Quality).

Harvey, L., with Geall, V., Mazelan, P., Moon, S. & Plummer, L. (1995) *Student satisfaction: The 1995 report on the student experience at UCE* (Birmingham, University of Central England, Centre for Research into Quality).

Harvey, L., with Plimmer, L., Moon, S. & Geall, V. (1997) *Student satisfaction manual* (Buckingham, SRHE & Open University Press).

Hativa, N. (1996) University instructors' ratings profiles: stability over time, and disciplinary differences, *Research in Higher Education*, 37, 341–365.

Hendry, G. D., Cumming, R. G., Lyon, P. M. & Gordon, J. (2001) Student-centred course evaluation in a four-year, problem based medical programme: issues in collection and management of feedback, *Assessment and Evaluation in Higher Education*, 26, 327–339.

Hennessy, S., Flude, M. & Tait, A. (1999) *An investigation of students' and tutors' views on tutorial provision: overall findings of the RTS project (phases I and II)* (Milton Keynes, The Open University, School of Education).

Husbands, C. T. & Fosh, P. (1993) Students' evaluation of teaching in higher education: experiences from four European countries and some implications of the practice, *Assessment and Evaluation in Higher Education*, 18, 95–114.

Information on quality and standards in higher education: final report of the task group (2002) (Report No. 02/15) (Bristol, Higher Education Funding Council for England).

Johnson, T. (1997) *The 1996 Course Experience Questionnaire: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).

Johnson, T. (1998) *The 1997 Course Experience Questionnaire: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).

Johnson, T. (1999) *Course Experience Questionnaire 1998: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).

Johnson, T., Ainley, J. & Long, M. (1996) *The 1995 Course Experience Questionnaire: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).

Kahl, T. N. & Cropley, A. J. (1999) Face-to-face versus distance learning: psychological consequences and practical implications, *Distance Education*, 7, 38–48.

Kember, D., Leung, D. Y. P. & 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.

Kidder, L. H. (1981) *Research methods in social relations* (New York, Holt, Rinehart & Winston).
Kolitch, E. & 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.
Kreber, C. (2003) The relationship between students' course perception and their approaches to studying in undergraduate science courses: a Canadian experience, *Higher Education Research and Development*, 22, 57–75.
Lawless, C. J. & Richardson, J. T. E. (2002) Approaches to studying and perceptions of academic quality in distance education, *Higher Education*, 44, 257–282.
Linke, R. D. (1991) *Performance indicators in higher education: report of a trial evaluation study commissioned by the Commonwealth Department of Employment, Education and Training* (Canberra, Australian Government Publishing Service).
Lizzio, A., Wilson, K. & Simons, R. (2002) University students' perceptions of the learning environment and academic outcomes: implications for theory and practice, *Studies in Higher Education*, 27, 27–52.
Long, M. & Hillman, K. (2000) *Course Experience Questionnaire 1999: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).
Lucas, L., Gibbs, G., Hughes, S., Jones, O. & Wisker, G. (1997) A study of the effects of course design features on student learning in large classes at three institutions: a comparative study, in: C. Rust & G. Gibbs (Eds) *Improving student learning through course design* (Oxford, Oxford Centre for Staff and Learning Development).
Lyon, P. M. & 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*, 27, 339–352.
Marsh, H. W. (1981) Students' evaluations of tertiary instruction: testing the applicability of American surveys in an Australian setting, *Australian Journal of Education*, 25, 177–192.
Marsh, H. W. (1982) SEEQ: a reliable, valid and useful instrument for collecting students' evaluations of university teaching, *British Journal of Educational Psychology*, 52, 77–95.
Marsh, H. W. (1983) Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics, *Journal of Educational Psychology*, 75, 150–166.
Marsh, H. W. (1986) Applicability paradigm: students' evaluations of teaching effectiveness in different countries, *Journal of Educational Psychology*, 78, 465–473.
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. W. (1991) Multidimensional students' evaluations of teaching effectiveness: a test of alternative higher-order structures, *Journal of Educational Psychology*, 83, 285–296.
Marsh, H. W. & Bailey, M. (1993) Multidimensional students' evaluations of teaching effectiveness, *Journal of Higher Education*, 64, 1–18.
Marsh, H. W. & Dunkin, M. J. (1992) Students' evaluations of university teaching: a multidimensional perspective, in: J. C. Smart (Ed.) *Higher education: handbook of theory and research, volume 8* (New York, Agathon Press).
Marsh, H. W. & Hocevar, D. (1991a) Multidimensional perspective on students' evaluations of teaching effectiveness: the generality of factor structures across academic discipline, instructor level, and course level, *Teaching and Teacher Education*, 7, 9–18.
Marsh, H. W. & Hocevar, D. (1991b) Students' evaluations of teaching effectiveness: the stability of mean ratings of the same teachers over a 13-year period, *Teaching and Teacher Education*, 7, 303–314.
Marsh, H. W. & Roche, L. A. (1992) The use of student evaluations of university teaching in different settings: the applicability paradigm, *Australian Journal of Education*, 36, 278–300.
Marsh, H. W. & Roche, L. A. (1994) *The use of students' evaluations of university teaching to improve teaching effectiveness* (Canberra, Department of Employment, Education and

Training). Available online at http://fistserv.uws.edu.au/seeq/Report/seeq1to5.htm (accessed 28 November 2002).

Marsh, H. W. & Roche, L. A. (1997) Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility, *American Psychologist*, 52, 1187–1197.

Marsh, H. W., Rowe, K. J. & Martin, A. (2002) PhD students' evaluations of research supervision: issues, complexities, and challenges in a nationwide Australian experiment in benchmarking universities, *Journal of Higher Education*, 73, 313–348.

Marsh, H. W., Touron, J. & Wheeler, B. (1985) Students' evaluations of university instructors: the applicability of American instruments in a Spanish setting, *Teaching and Teacher Education*, 1, 123–138.

Mazuro, C., Norton, L., Hartley, J., Newstead, S. & Richardson, J.T.E. (2000) Practising what you preach? Lecturers' and students' perceptions of teaching practices, *Psychology Teaching Review*, 9, 91–102.

McInnis, C., Griffin, P., James, R. & Coates, H. (2001) *Development of the Course Experience Questionnaire (CEQ)* (Canberra, Department of Education, Training and Youth Affairs).

Meyer, J. H. F. & Muller, M. W. (1990) Evaluating the quality of student learning: I. An unfolding analysis of the association between perceptions of learning context and approaches to studying at an individual level, *Studies in Higher Education*, 15, 131–154.

Moore, M. G. (1980) Independent study, in: R. D. Boyd, J. W. Apps & Associates, *Redefining the discipline of adult education* (San Francisco, Jossey-Bass).

Morgan, A. (1984) A report on qualitative methodologies in research in distance education, *Distance Education*, 5, 252–267.

Murray, H. G. (1983) Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness, *Journal of Educational Psychology*, 75, 138–149.

Narasimhan, K. (2001) Improving the climate of teaching sessions: the use of evaluations by students and instructors, *Quality in Higher Education*, 7, 179–190.

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

Neumann, R. (2000) Communicating student evaluation of teaching results: rating interpretation guides (RIGs), *Assessment and Evaluation in Higher Education*, 25, 121–134.

Nielsen, H. D., Moos, R. H. & Lee, E. A. (1978) Response bias in follow-up studies of college students, *Research in Higher Education*, 9, 97–113.

Parsons, P. G. (1988) The Lancaster approaches to studying inventory and course perceptions questionnaire: a replicated study at the Cape Technikon, *South African Journal of Higher Education*, 2, p103–111.

Pearson, M., Kayrooz, C. & Collins, R. (2002) Postgraduate student feedback on research supervisory practice, paper presented at the *Annual Conference of the Society for Research into Higher Education*, University of Glasgow.

Perry, W. G. (1970) *Forms of intellectual and ethical development in the college years: a scheme* (New York, Holt, Rinehart & Winston).

Pozo-Muñoz, C., Rebolloso-Pacheco, E. & Fernández-Ramirez, B. (2000) The 'ideal teacher': implications for student evaluation of teacher effectiveness, *Assessment and Evaluation in Higher Education*, 25, 253–263.

Prosser, M., Trigwell, K., Hazel, E. & Gallagher, P. (1994) Students' experiences of teaching and learning at the topic level, *Research and Development in Higher Education*, 16, 305–310.

Purcell, K. & Pitcher, J. (1998) *Great expectations: a new diversity of graduate skills and aspirations* (Manchester, Higher Education Careers Services Unit).

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

Ramsden, P. (1991b) Report on the Course Experience Questionnaire trial, in: R. D. Linke *Performance indicators in higher education: report of a trial evaluation study commissioned by the*