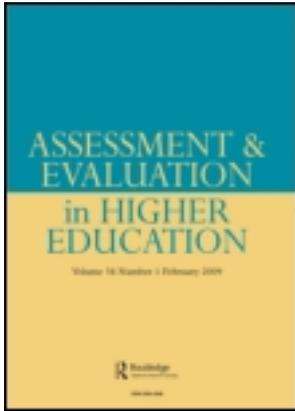


This article was downloaded by: [JAMES COOK UNIVERSITY]

On: 17 January 2013, At: 14:29

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caeh20>

Instruments for obtaining student feedback: a review of the literature

John T. E. Richardson^a

^a The Open University, UK

Version of record first published: 14 Sep 2010.

To cite this article: John T. E. Richardson (2005): Instruments for obtaining student feedback: a review of the literature, *Assessment & Evaluation in Higher Education*, 30:4, 387-415

To link to this article: <http://dx.doi.org/10.1080/02602930500099193>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Instruments for obtaining student feedback: a review of the literature

John T. E. Richardson*

The Open University, UK

This paper reviews the research evidence concerning the use of formal instruments to measure students' evaluations of their teachers, students' satisfaction with their programmes and students' perceptions of the quality of their programmes. These questionnaires can provide important evidence for assessing the quality of teaching, for supporting attempts to improve the quality of teaching and for informing prospective students about the quality of course units and programmes. The paper concludes by discussing several issues affecting the practical utility of the instruments that can be used to obtain student feedback. Many students and teachers believe that student feedback is useful and informative, but for a number of reasons many teachers and institutions do not take student feedback sufficiently seriously.

Introduction

The purpose of this article is to review the published research literature concerning the use of formal instruments to obtain student feedback in higher education. My primary emphasis will be on sources that have been subjected to the formal processes of independent peer review, but there is also a 'grey' literature consisting of conference proceedings, in-house publications and technical reports that contain relevant information even if they are lacking in academic rigour.

The first part of my review will cover the predominantly North American literature that is concerned with students' evaluations of their teachers. I shall briefly refer to attempts to measure student satisfaction and then turn to the predominantly Australian and British literature that is concerned with students' perceptions of the quality of their programmes. The final part of my review will deal with more practical issues: why collect student feedback? Why use formal instruments? What should be the subject of the feedback? What kind of feedback should be collected? When should feedback be collected? Would a single questionnaire be suitable for all

*Corresponding author. Institute of Educational Technology, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. Email: J.T.E.Richardson@open.ac.uk

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.
- Information for students to use in the selection of course units and teachers.
- An outcome or process description for use in research on teaching.

Marsh and Dunkin noted that the first purpose was essentially universal in North America, but the other three were not:

At many universities systematic student input is required before faculty are even considered for promotion, while at others the inclusion of SETs is optional or not encouraged at all. Similarly, in some universities the results of SETs are sold to students in university bookstores as an aid to the selection of courses or instructors, whereas the results are considered to be strictly confidential at other universities. (1992, p. 143)

The feedback in question usually takes the form of students' ratings of their level of satisfaction or their self-reports of other attitudes towards their teachers or their course units. The feedback is obtained by means of standard questionnaires, the responses are automatically scanned, and a descriptive summary of the responses is returned to the relevant teacher and, if appropriate, the teacher's head of department. The process is relatively swift, simple and convenient for both students and teachers, and in most North American institutions it appears to have been accepted as a matter of routine. It has, however, been described as a 'ritual' (Abrami *et al.*, 1996, p. 213), and precisely for that reason it may not always be regarded as a serious matter by those involved. In many institutions, the instruments used to obtain student feedback have been constructed and developed in-house and may never have been subjected to any kind of external scrutiny. Marsh (1987) described five instruments that had received some kind of formal evaluation and others have featured in subsequent research.

The instrument that has been most widely used in published work is Marsh's (1982) Students' Evaluations of Educational Quality (SEEQ). In completing this questionnaire, students are asked to judge how well each of 35 statements (for instance, 'You found the course intellectually stimulating and challenging') describes their teacher or course unit, using a five-point scale from 'very poor' to 'very good'. The statements are intended to reflect nine aspects of effective teaching: learning/value, enthusiasm, organization, group interaction, individual rapport, breadth of coverage, examinations/grading, assignments and workload/difficulty. The evidence using this and other similar questionnaires has been summarized in a series of reviews (Marsh, 1982, 1987; Arubayi, 1987; Marsh & Dunkin, 1992; Marsh & Bailey, 1993).

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 highly stable over time. Indeed, Marsh and Hocevar (1991b) found no systematic changes in students' ratings of 195 teachers over a 13-year period. Although this demonstrates the stability of the students' ratings, it also implies that the performance of the teachers was not improving with experience. Nevertheless, Roche and Marsh (2002) found that teachers' perceptions of their own teaching became more consistent with their students' perceptions of their teaching as a result of receiving feedback in the form of students' evaluations. In other words, students' evaluations may change teachers' self-perceptions even if they do not change their teaching behaviour.

The factor structure of the SEEQ has been confirmed in several studies. In particular, Marsh and Hocevar (1991a) showed that it was invariant across teachers of different status and across course units in different disciplines and at different levels. There is a consensus that students' ratings of teaching effectiveness vary on a large number of dimensions, but there is debate as to whether these can be subsumed under a single, more global dimension. Marsh (1991; Marsh & Dunkin, 1992; Marsh & Roche, 1997) argued that, although students' scores on the dimensions of the SEEQ were correlated with each other, they could not be adequately captured by a single higher-order factor. On the other hand, Abrami and d'Apollonia (1991; Abrami *et al.*, 1996; d'Apollonia & Abrami, 1997) proposed that students' evaluations of teaching were subsumed by a single overarching construct that they defined as 'general instructional skill'.

The fact that students' evaluations of teachers are correlated with the teachers' self-evaluations also constitutes evidence for their validity. In fact, teachers' self-evaluations exhibit essentially the same factor structure as their students' evaluations, teachers' self-evaluations are correlated with their students' evaluations on each individual dimension of the SEEQ, and teachers' self-evaluations are not systematically different from their students' evaluations (see Marsh, 1987). Students' evaluations of their teachers are not highly correlated with evaluations provided by *other* teachers on the basis of classroom observation, but both the reliability and the validity of the latter evaluations have been questioned. There is better evidence that SETs are correlated with ratings of specific aspects of teaching by trained observers (see Murray, 1983).

In principle, the validity of students' evaluations might be demonstrated by finding correlations between SETs and academic performance. However, the demands and the assessment criteria of different course units may vary, and so students' grades or examination marks cannot be taken as a simple measure of teaching effectiveness. One solution is to compare students' evaluations and attainment in a single course

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 average SETs and the average attainment across different groups, and even when the different groups of students do not vary significantly in terms of the grades that they expect to achieve. Marsh argued that this was inconsistent with the grading satisfaction hypothesis and supported the validity of SETs.

Although the SEEQ has been most widely used in North America, it has also been employed in investigations carried out in Australia, New Zealand, Papua New Guinea and Spain (Marsh, 1981, 1986; Clarkson, 1984; Marsh *et al.*, 1985; Watkins *et al.*, 1987; Marsh & Roche, 1992). The instrument clearly has to be adapted (or

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-development package for university teachers that incorporates a self-rating questionnaire for teachers, a guide to interpreting the students' overall evaluations, and booklets on improving teaching effectiveness in areas where evaluations identify scope for improvement. They offered advice on how this package might be adopted in programmes at other institutions.

Marsh (1987) concluded that 'student ratings are clearly multidimensional, quite reliable, reasonably valid, relatively uncontaminated by many variables often seen as sources of potential bias, and are seen to be useful by students, faculty, and administrators' (p. 369). The literature that has been published over the subsequent period has confirmed each of these points and has also demonstrated that student ratings can provide important evidence for research on teaching. The routine collection of students' evaluations does not in itself lead to any improvement in the quality of teaching (Kember *et al.*, 2002). Nevertheless, feedback of this nature may help in the professional development of individual teachers, particularly if it is supported by an appropriate process of consultation and counselling (Roche & Marsh, 2002). SETs do increase systematically following specific interventions aimed at improving teaching (Hativa, 1996).

Student satisfaction surveys

Perhaps the most serious limitation of the instruments that have been described thus far is that they have focused upon students' evaluations of particular course units in the context of highly modular programmes of study, and hence they provide little information about their experience of their programmes or institutions as a whole. In addition to collecting SETs for individual course units, many institutions in North America make use of commercially published questionnaires to collect comparative data on their students' overall satisfaction as consumers.

One widely used questionnaire is the Noel-Levitz Student Satisfaction Inventory, which is based explicitly on consumer theory and measures students' satisfaction with their experience of higher education. It contains either 76 items (for institutions offering two-year programmes) or 79 items (for institutions offering four-year programmes); in each case, respondents are asked to rate both the importance of their expectation about a particular aspect of higher education and their level of satisfaction. Overall scores are calculated that identify aspects of the students' experience where the institutions are failing to meet their expectations.

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 raised about the adequacy of the CPQ as a research tool (Meyer & Muller, 1990).

Ramsden (1991a) developed a revised instrument, the Course Experience Questionnaire (CEQ), as a performance indicator for monitoring the quality of teaching

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 items of the five scales (according to the results of the national trial) are shown in Table 1. In addition, three of the items in the Appropriate Assessment scale could be used as a subscale to monitor the perceived importance of rote memory as opposed to understanding in assessment.

The respondents were instructed to indicate their level of agreement or disagreement (along a scale from 'definitely agree', scoring five, to 'definitely disagree', scoring one) with each statement as a description of their programme of study. Half of the items referred to positive aspects, whereas the other half referred to negative aspects and were to be scored in reverse. This means that the instrument as a whole controlled for any systematic responses biases either to agree with all of the items or to disagree with all of the items. (Unfortunately, the items to be scored in reverse were not distributed equally across the five CEQ scales.)

As a result of this national trial, it was determined that the Graduate Careers Council of Australia (GCCA) should administer the CEQ on an annual basis to all new graduates through the Graduate Destination Survey, which is conducted a few months after the completion of their degree programmes. The survey of the 1992 graduates was carried out in 1993 and obtained usable responses to the CEQ from more than 50,000 graduates from 30 institutions of higher education (Ainley & Long, 1994). Subsequent surveys have covered all Australian universities and have typically obtained usable responses to the CEQ from more than 80,000 graduates, reflecting

Table 1. Defining items of the scales in the original Course Experience Questionnaire

Scale	Defining item
Good teaching	Teaching staff here normally give helpful feedback on how you are going.
Clear goals and standards	You usually have a clear idea of where you're going and what's expected of you in this course.
Appropriate workload	The sheer volume of work to be got through in this course means you can't comprehend it all thoroughly.
Appropriate assessment	Staff here seem more interested in testing what we have memorized than what we have understood.
Emphasis on independence	Students here are given a lot of choice in the work they have to do.

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 length of the questionnaire still further, the Emphasis on Independence scale was dropped, and a further seven items were removed on the grounds that they had shown only a weak relationship with the scales to which they had been assigned in Ramsden’s (1991a, b, p. 6) analysis of the data from the Australian national trial. This produced a revised, short form of the CEQ consisting of 23 items in five scales.
- Two other items were employed but not assigned to any of the scales. One measured the respondents’ overall level of satisfaction with their programmes, and this has proved to be helpful in validating the CEQ as an index of perceived academic quality (see below). An additional item in the first two surveys was concerned with the extent to which respondents perceived their programmes to be overly theoretical or abstract. This was replaced in the next three surveys by reinstating an item from the Appropriate Assessment scale that measured the extent to which feedback on their work was usually provided only in the form of marks or grades. In subsequent surveys, this in turn was replaced by a wholly new item concerned with whether the assessment methods required an in-depth understanding of the syllabus. In practice, however, the responses to these additional items have not shown a strong relationship with those given to other items from the Appropriate Assessment scale, and so they have not been used in computing the respondents’ scale scores.

Wilson *et al.* (1997) proposed that for research purposes the original version of the CEQ should be augmented with the Generic Skills scale to yield a 36-item instrument. They compared the findings obtained using the short, 23-item version and this 36-item version when administered to successive cohorts of graduates from one Australian university.

Evidence concerning the psychometric properties of the 30-item version of the CEQ has been obtained in the Australian national trial (Ramsden, 1991a, b) and in

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 students taking different course units and their self-reported approaches to studying, such that the students who evaluate their course units more positively on the CEQ are more likely to adopt a deep approach to learning (Lawless & Richardson, 2002; Richardson, 2003, in press; Richardson & Price, 2003). Typically, the two sets of measures share between 45% and 80% of their variance. Similar results were obtained by Trigwell and Ashwin (2002), who used an adapted version of the CEQ to assess perceptions of the tutorial system among students at an Oxford college, and by Sadlo and Richardson (2003), who used the original CEQ to assess perceptions of academic quality among students taking subject-based and problem-based programmes in six different schools of occupational therapy.

Wilson *et al.* (1997) demonstrated that students' scores on the 36-item version of the CEQ were significantly correlated with their cumulative grade point averages. The correlation coefficients were highest for the Good Teaching scale and the Clear Goals and Standards scale, and they were lowest for the Generic Skills and Appropriate Workload scales. Of course, these data do not imply a causal link between good teaching and better grades. As mentioned above, Marsh (1987) pointed out that more positive student ratings could result from students' satisfaction at receiving higher grades or from uncontrolled characteristics of the student population. In both these cases, however, it is not clear why the magnitude of the relationship between CEQ scores and academic attainment should vary across different scales of the CEQ.

Finally, Lizzio *et al.* (2002) constructed a theoretical model of the relationships between CEQ scores, approaches to studying and academic outcomes. They interpreted scores on the Generic Skills scale and students' overall ratings of satisfaction as outcome measures, as well as grade point average. In general, they found that students' scores on the other five scales of the CEQ were positively correlated with all three outcome measures. Students' perceptions of their academic environment according to the CEQ had both a direct influence upon academic outcomes and an indirect influence that was mediated by changes in the students' approaches to studying. In contrast, students' academic achievement before their admission to university had only a weak influence on their grade point average and no effect on their overall satisfaction.

Although the CEQ has been predominantly used in Australia, it has also been used in other countries to compare graduates and current students from different programmes. For instance, Sadlo (1997) used the CEQ to compare students taking

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 each programme at each institution is published, provided that the response rate at the institution in question has exceeded 50%. (Normally this is achieved by all except a few private institutions.) Once again, the process seems to have been accepted as a matter of routine in most Australian institutions, and some are using versions of the CEQ to monitor their current students, too. At the University of Sydney, for example, the average ratings on an adapted version of the CEQ determine a portion of the financial resources allocated to each faculty.

One problem is that the wording of the individual items in the CEQ may not be suitable for all students. For instance, Johnson *et al.* (1996, p. 3) remarked that the appropriateness of some items was questionable in the case of respondents who had completed a qualification through a programme of research, since in this case the notion of meeting the requirements of a particular 'course' might be quite tenuous.

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 increasingly 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 CEQ to measure students' expectations when embarking on a programme in addition to their subsequent perceptions of its academic quality.) This theoretical approach was extended by Narasimhan (2001) to include the expectations and perceptions of teachers in higher education as well as those of their students.

One fundamental difficulty with this approach is that it privileges satisfaction as a notion that is coherent, homogeneous and unproblematic. In fact, the limited amount of research on this topic suggests that student satisfaction is a complex yet poorly articulated idea that is influenced by a wide variety of contextual factors that are not intrinsically related to the quality of teaching (Wiers-Jenssen *et al.*, 2002). On theoretical grounds, it is not at all clear that satisfaction should be a desirable outcome of higher education, let alone that it should be likened to a commodity or service. Indeed, the discomfort that is associated with genuine intellectual growth has been well documented in interview-based research by Perry (1970) and Baxter Magolda (1992). In the case of research using the CEQ, in contrast, students' ratings of overall satisfaction with their courses or programmes are simply used as a way of validating students' perceptions of academic quality.

A different issue is whether student feedback should be concerned solely with curricular matters or whether it should also be concerned with the various facilities available at institutions of higher education (including computing, library, recreational and sporting facilities). It cannot be denied that the latter considerations are important in the wider student experience. However, students' ratings of these facilities are not highly correlated with their perceptions of the quality of teaching and learning (Yorke, 1995), and they are less important as predictors of their overall satisfaction than their perceptions of the academic features of their programmes (Wilson *et al.*, 1997). As noted in the case of the CEQ, including additional scales about the wider institutional environment might tend to undermine feedback questionnaires as indicators of teaching quality. It would be preferable to evaluate institutional facilities as an entirely separate exercise, and in this case an approach orientated towards consumer satisfaction might well be appropriate.

When should feedback be collected?

It would seem sensible to collect feedback on students' experience of a particular educational activity at the completion of that activity, since it is presumably their experience of the entire activity that is of interest. In other words, it would be most appropriate to seek student feedback at the end of a particular course unit or programme of study. However, other suggestions have been made. Narasimhan (2001) noted that obtaining feedback at the end of a course unit could not benefit 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 programmes. This would be a serious possibility if future questionnaires used for the FDS were perceived as cumbersome or intrusive.

Would a single questionnaire be suitable for all students?

Experience with the SEEQ and the CEQ in America and Australia suggests that it is feasible to construct questionnaires that have a very wide range of applicability. The results have been used to make meaningful comparisons across a variety of institutions and a variety of disciplines. In addition, many institutions that use the SEEQ to obtain feedback from students about teachers or course units, and many institutions that use the CEQ to obtain feedback from recent graduates about their programmes seem to accept these surveys as sufficient sources of information and do not attempt to supplement them with other instruments. This suggests that the introduction of a single national questionnaire to survey recent graduates (see below) might supplant instruments that are currently used for this purpose by individual institutions. Instead, they might be induced to focus their efforts elsewhere (for instance, in more extensive surveys of current students).

It is clearly necessary that such a questionnaire should be motivated by research evidence about teaching, learning and assessment in higher education and that it should be assessed as a research tool. The only existing instruments that satisfy these requirements are the SEEQ (for evaluating individual teachers and course units) and the CEQ (for evaluating programmes). It has been argued that instruments like the SEEQ take for granted a didactic model of teaching, and this may be true of any questionnaire that focuses on the role of the teacher at the expense of the learner. Conversely, course designers who adopt more student-focused models of teaching may find that these instruments are unhelpful as evaluative tools (Kember *et al.*, 2002). In a similar way, Lyon and Hendry (2002) claimed that the CEQ was not appropriate for evaluating programmes with problem-based curricula. However, their results may have been due not to inadequacies of the CEQ but to difficulties that they encountered in introducing problem-based learning (Hendry *et al.*, 2001). The CEQ

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

demographic biases. However, respondents differ from nonrespondents in their attitudes and behaviour in ways that cannot be predicted on the basis of known demographic characteristics (Goyder, 1987, chapter 7). In particular, students who respond to surveys differ from those who do not respond in terms of their study behaviour and academic attainment (Astin, 1970; Nielsen *et al.*, 1978; Watkins & Hattie, 1985). It is therefore reasonable to assume that students who respond to feedback questionnaires will be systematically different from those who do not respond in their attitudes and experience of higher education. This kind of bias is unavoidable, and it cannot be addressed by a simple weighting strategy. Nevertheless, its impact can be reduced by minimizing the number of non-respondents.

In social research, a response rate of 50% is considered satisfactory for a postal survey (Babbie, 1973, p. 165; Kidder, 1981, pp. 150–151). As mentioned earlier, the Australian GCCA surveys require that this response rate be achieved by individual institutions if their average ratings are to be published. Indeed, the vast majority of participating institutions do achieve this response rate, and at a national level the GCCA surveys regularly achieve response rates of around 60%. In other words, this is the kind of response rate that can be achieved in a well-designed postal survey, although it clearly leaves ample opportunity for sampling bias to affect the results. The position of the Australian Vice-Chancellor's Committee (2001) is that an overall institutional response rate for the CEQ of at least 70% is both desirable and achievable.

Student feedback at the end of course units is often collected in a class situation, and this could be used to obtain feedback at the end of entire programmes (in both cases, presumably, before the assessment results are known). This is likely to yield much higher response rates and hence to reduce the impact of sampling error and sampling bias. There is an ethical issue as to whether students should be required to contribute feedback in this manner. In a class situation, students might feel under pressure to participate in the process, but the guidelines of many professional bodies stipulate that participants should be able to withdraw from a research study at any time. It will be important for institutions to clarify whether the collection of feedback is a formal part of the teaching-learning process or whether it is simply tantamount to institutional research.

With the increasing use of information technology in higher education, institutions may rely less on classroom teaching and more upon electronic forms of communication. This is already the case in distance learning, where electronic means of course delivery are rapidly replacing more traditional correspondence methods. Information technology can also provide a very effective method of administering social surveys, including the direct electronic recording of responses (see Watt *et al.*, 2002). It would be sensible to administer feedback surveys by the same mode as that used for delivering the curriculum (classroom administration for face-to-face teaching, postal surveys for correspondence courses and electronic surveys for on-line courses). Little is known about the response rates obtained in electronic surveys, or whether different modes of administration yield similar patterns of results. It is however good practice to make feedback questionnaires available in a variety of formats for use by students

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 competence contribute to the idea of an 'ideal teacher' (Pozo-Muñoz *et al.*, 2000), but most teachers do consider SETs to be useful sources of information (Schmelkin *et al.*, 1997). Left to their own devices, however, they may be unlikely to change their teaching in the light of the results, to make the results available for other students, to discuss them with more senior members of staff or to refer them to institutional committees or administrators (Nasser & Fresko, 2002).

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 National Student Survey. This was carried out during the summer of 2003 using an inventory containing 48 items and obtained responses from 17,173 students at 22 institutions. Subsequently, however, the UK Government decided that it would be more efficient to survey students during, rather than after, their final year of study. A second pilot survey was therefore administered to final-year students at ten institutions early in 2004; this used a 37-item inventory and yielded responses from 9723 students. The results from these pilot surveys will inform a full survey of all final-year students in England, Wales and Northern Ireland to be carried out early in 2005. The findings will not be reported here because they have yet to be subjected to peer review, but preliminary accounts can be found at the National Student Survey website: <http://iet.open.ac.uk/nss>.

What can be said, in the meantime, is that the experience of these pilot surveys serves to confirm many of the points made earlier in this article. More generally, the published research literature leads one to the following conclusions:

- Student feedback provides important evidence for assessing quality, it can be used to support attempts to improve quality, and it can be useful to prospective students.
- The use of quantitative instruments is dictated by organizational constraints (and in distance education by geographical constraints, too).
- Feedback should be sought at the level at which one is endeavouring to monitor quality.
- The focus should be on students' perceptions of key aspects of teaching or on key aspects of the quality of their programmes.
- Feedback should be collected as soon as possible after the relevant educational activity.
- It is feasible to construct questionnaires with a very wide range of applicability. Two groups are problematic: postgraduate research students and distance-learning students. Curricular innovations might make it necessary to reword or more radically amend existing instruments. In addition, any comparisons among different course units or programmes should take into account the diversity of educational contexts and student populations.
- Response rates of 60% or more are both desirable and achievable for students who have satisfactorily completed their course units or programmes. Response rates

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 at the UK Open University. He is the author of *Researching student learning: approaches to studying in campus-based and distance education* (Buckingham, SRHE & Open University Press, 2000).

References

- Abrami, P. C. & d'Apollonia, S. (1991) Multidimensional students' evaluations of teaching effectiveness—generalizability of 'N = 1' research: comment on Marsh (1991), *Journal of Educational Psychology*, 83, 411–415.
- Abrami, P. C., d'Apollonia, S. & Rosenfield, S. (1996) The dimensionality of student ratings of instruction: what we know and what we do not, in: J. C. Smart (Ed.) *Higher education: handbook of theory and research*, volume 11 (New York, Agathon Press).
- Ainley, J. (1999) Using the Course Experience Questionnaire to draw inferences about higher education, paper presented at the conference of the *European Association for Research on Learning and Instruction*, Göteborg, Sweden.
- Ainley, J. & Long, M. (1994) *The Course Experience survey 1992 graduates* (Canberra, Australian Government Publishing Service).
- Ainley, J. & Long, M. (1995) *The 1994 Course Experience Questionnaire: a report prepared for the Graduate Careers Council of Australia* (Parkville, Victoria, Graduate Careers Council of Australia).
- Arubayi, E. A. (1987) Improvement of instruction and teacher effectiveness: are student ratings reliable and valid?, *Higher Education*, 16, 267–278.
- Astin, A. W. (1970) The methodology of research on college impact, part two, *Sociology of Education*, 43, 437–450.

- 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 Education*, 25, 221–236.
- Baxter Magolda, M. B. (1992) *Knowing and reasoning in college: gender-related patterns in students' intellectual development* (San Francisco, Jossey-Bass).
- Broomfield, D. & Bligh, J. (1998) An evaluation of the 'short form' Course Experience Questionnaire with medical students, *Medical Education*, 32, 367–369.
- Byrne, M. & Flood, B. (2003) Assessing the teaching quality of accounting programmes: an evaluation of the Course Experience Questionnaire, *Assessment and Evaluation in Higher Education*, 28, 135–145.
- Cheng, D. X. (2001) Assessing student collegiate experience: where do we begin?, *Assessment and Evaluation in Higher Education*, 26, 525–538.
- Chiang, K. H. (2002) Relationship between research and teaching in doctoral education in UK universities, paper presented at the *Annual Conference of the Society for Research into Higher Education*, University of Glasgow.
- Clarkson, P. C. (1984) Papua New Guinea students' perceptions of mathematics lecturers, *Journal of Educational Psychology*, 76, 1386–1395.
- Coffey, M. & Gibbs, G. (2000) Can academics benefit from training? Some preliminary evidence, *Teaching in Higher Education*, 5, 385–389.
- Coffey, M. & Gibbs, G. (2001) The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education, *Assessment and Evaluation in Higher Education*, 26, 89–93.
- Coffey, M. & Gibbs, G. (in press) New teachers' approaches to teaching and the utility of the Approaches to Teaching Inventory, *Higher Education Research and Development*.
- Cohen, P. A. (1981) Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies, *Review of Educational Research*, 51, 281–309.
- Committee of Vice-Chancellors and Principals (1998) *Skills Development in Higher Education: A Short Report* (London, Committee of Vice-Chancellors and Principals).
- Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests, *Psychometrika*, 16, 297–334.
- Curtin University of Technology Teaching Learning Group (1997) *Student evaluation of teaching at Curtin University: piloting the student evaluation of educational quality (SEEQ)* (Perth, WA, Curtin University of Technology).
- D'Apollonia, S. & Abrami, P. C. (1997) Navigating student ratings of instruction, *American Psychologist*, 52, 1198–1208.
- Eley, M. G. (1992) Differential adoption of study approaches within individual students, *Higher Education*, 23, 231–254.
- Fung, Y. & Carr, R. (2000) Face-to-face tutorials in a distance learning system: meeting student needs, *Open Learning*, 15, 35–46.
- Gibbs, G. & Coffey, M. (2001) *Developing an associate lecturer student feedback questionnaire: evidence and issues from the literature* (Student Support Research Group Report No. 1) (Milton Keynes, The Open University).
- Gibbs, G., Habeshaw, S. & Habeshaw, T. (1988) *53 interesting ways to appraise your teaching* (Bristol, Technical and Educational Services).

- 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., Reboloso-Pacheco, E. & Fernández-Ramírez, 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*

- Commonwealth Department of Employment, Education and Training. Volume 2: supplementary papers (Canberra, Australian Government Publishing Service).
- Ramsden, P. & Entwistle, N. J. (1981) Effects of academic departments on students' approaches to studying, *British Journal of Educational Psychology*, 51, 368–383.
- Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests* (Copenhagen, Denmark's pædagogiske Institut).
- Richardson, J. T. E. (1994) A British evaluation of the Course Experience Questionnaire, *Studies in Higher Education*, 19, 59–68.
- Richardson, J. T. E. (1997) Students' evaluations of teaching: the Course Experience Questionnaire, *Psychology Teaching Review*, 6, 31–45.
- Richardson, J. T. E. (2000) *Researching student learning: approaches to studying in campus-based and distance education* (Buckingham, SRHE and Open University Press).
- Richardson, J. T. E. (2003) Approaches to studying and perceptions of academic quality in a short web-based course, *British Journal of Educational Technology*, 34, 433–442.
- Richardson, J. T. E. (2005) Students' perceptions of academic quality and approaches to studying in distance education, *British Educational Research Journal*, 31(1), 1–21.
- Richardson, J. T. E. & Price, L. (2003) Approaches to studying and perceptions of academic quality in electronically delivered courses, *British Journal of Educational Technology*, 34, 45–56.
- Richardson, J. T. E. & Woodley, A. (2001) Perceptions of academic quality among students with a hearing loss in distance education, *Journal of Educational Psychology*, 93, 563–570.
- Roche, L. A. & Marsh, H. W. (2002) Teaching self-concept in higher education: reflecting on multiple dimensions of teaching effectiveness, in: N. Hativa & P. Goodyear (Eds) *Teacher thinking, beliefs and knowledge in higher education* (Dordrecht, Kluwer).
- Sadlo, G. (1997) Problem-based learning enhances the educational experiences of occupational therapy students, *Education for Health*, 10, 101–114.
- Sadlo, G. & Richardson, J. T. E. (2003). Approaches to studying and perceptions of the academic environment in students following problem-based and subject-based curricula, *Higher Education Research and Development*, 22, 253–274.
- Schmelkin, L. P., Spencer, K. J. & Gillman, E. S. (1997) Faculty perspectives on course and teacher evaluations, *Research in Higher Education*, 38, 575–592.
- Spencer, K. J. & Schmelkin, L. P. (2002) Student perspectives on teaching and its evaluation, *Assessment and Evaluation in Higher Education*, 27, 397–409.
- Trigwell, K. & Ashwin, P. (2002) Evoked conceptions of learning and learning environments, paper presented at the *Tenth International Improving Student Learning Symposium*, Brussels.
- Trigwell, K. & Prosser, M. (1991) Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes, *Higher Education*, 22, 251–266.
- Watkins, D. & Hattie, J. (1985) A longitudinal study of the approaches to learning of Australian tertiary students, *Human Learning*, 4, 127–141.
- Watkins, D., Marsh, H. W. & Young, D. (1987) Evaluating tertiary teaching: a New Zealand perspective, *Teaching and Teacher Education*, 2, 41–53.
- Watt, S., Simpson, C., McKillop, C. & Nunn, V. (2002) Electronic course surveys: does automating feedback and reporting give better results? *Assessment and Evaluation in Higher Education*, 27, 325–337.
- Waugh, R. F. (1998) The Course Experience Questionnaire: a Rasch measurement model analysis, *Higher Education Research and Development*, 17, 45–64.
- Wiers-Jenssen, J., Stensaker, B. & Grøgaard, J. B. (2002) Student satisfaction: towards an empirical deconstruction of the concept, *Quality in Higher Education*, 8, 183–195.
- Wilson, K. L., Lizzio, A. & Ramsden, P. (1997) The development, validation and application of the Course Experience Questionnaire, *Studies in Higher Education*, 22, 33–53.
- Yorke, M. (1995) Siamese twins? Performance indicators in the service of accountability and enhancement, *Quality in Higher Education*, 1, 13–30.