The Development of an Extended Course Experience Questionnaire

PATRICK GRIFFIN , HAMISH COATES , CRAIG MCINNIS & RICHARD JAMES

The Development of an Extended Course Experience Questionnaire

PATRICK GRIFFIN, HAMISH COATES, CRAIG MCINNIS & RICHARD JAMES

The University of Melbourne, Australia

ABSTRACT *Universities in Australia have been using the course experience questionnaire (CEQ) for a number of years to measure the quality of teaching. Recently, there have been increasing calls for a broader perspective that takes into account the potential of non-classroom context influences on the student learning experience. This project undertook to develop a broader instrument that added a range of scales that could be linked to the existing instrument. A sample of almost 4000 students responded to the trial instrument. The trials suggest that the existing 'good teaching', 'generic skills', 'clear goals', 'appropriate workload' and 'appropriate assessment' scales can be supplemented by the following additional scales: 'student support', 'learning resources', 'learning community', 'graduate qualities' and 'intellectual motivation'. These new scales were shown through Rasch analyses to be psychometrically reliable and accurate.*

The Development of an Extended Course Experience Questionnaire

There has been a range of methods proposed for evaluating the quality of undergraduate programmes. Research by Marsh (1987) and Ramsden (1991a) suggests that student evaluations are among the most valid and reliable. Despite their shortcomings and potential for misuse, surveys of student perceptions now play a significant role in the higher education sector. An instance of this is the use of graduate evaluation data for comparative purposes in the *Good Universities Guide* (Ashenden & Milligan, 2000). Given the increasing and widespread use of student evaluations, there is a need to monitor and evaluate their effectiveness as tools in the quality management process and, where possible, take action to ensure they provide valid and reliable information about institutional performance.

Since 1993 the course experience questionnaire (CEQ) has been mailed to every person completing an undergraduate qualification in Australia in the year following completion of their course. The CEQ is intended to probe key elements of the university learning process and, in doing so, obtain data on the quality of teaching and courses. Rather than measuring the multitude of factors that combine to form student experience, the CEQ was developed with an assumption of a strong association between the quality of student learning and student perceptions of teaching. The items and scales are specifically tuned to obtain information on what were considered by Ramsden (1991a, 1991b) to be the defining elements of teaching and its organisation. By considering the extent to which instruction encourages deep, rather than surface, understanding of concepts and materials, the CEQ attempts to provide a domain-neutral indicator of university course quality (McInnis, 1997).

The CEQ was developed for two principal reasons. First, the instrument was intended

to facilitate quality assurance and accountability. The primary and motivating purpose of the instrument is to use student ratings to derive performance indicators of teaching effectiveness in higher education institutions. As a 'direct measure of consumer satisfaction with higher education', the CEQ was developed to produce 'as economically as possible … ordinal ranking of academic organisational units in different institutions' (Ramsden, 1991a, p. 130; 1991b). Second, by internal and comparative evaluation of results, the CEQ outputs are intended to assist institutions with their quality enhancement and improvement processes. Performance indicators have the potential to drive performance, as well as measure it. Analysis of information from ex-students initiates a feedback loop capable of providing valuable information to institutions on the benefits and constraints of particular courses. A market-driven perspective further suggests that comparative evaluation will promote competition and lead to institutional improvements. Although distribution of sensitive performance data may seem to contradict internal commercial objectives, in an increasingly competitive national system there may be mutual benefits if benchmarking activities can objectively reveal areas of weakness and point to mechanisms for improvement. Co-operation and collaboration may motivate more effective competition and consequent quality improvement. A third, largely *ex-post facto* series of functions of the instrument are more student-oriented. In addition to the standardised quantitative results of the evaluation process being used by institutions for marketing purposes, they have also been incorporated into more commercial texts such as the *Good Universities Guide* (Ashenden & Milligan, 2000). Additionally, and importantly, the survey process provides graduates with a framework and formal means of reflecting on their courses.

It is clear, given these purposes, that the CEQ must provide information that is both nationally generalisable and locally sensitive. The instrument needs to possess qualities that enable national benchmarking across fields of study, are sufficiently detailed to facilitate accurate management of each institution's quality-improvement processes, and provide students with a response framework relevant to their course. The items and scales need to mediate between abstraction and sensitivity and, without fostering an institutional relativism, the instrument needs to provide data that adequately represents the 'lifeworld' of each institution. If quality assurance or management processes are built upon unrepresentative information, innovation and diversity could be hindered. As McInnis (1997, p. 65) argues, 'universities are inevitably tempted to comply with, or at least drift towards' the parameters set by performance indicators.

The original CEQ was based on a theory of learning that emphasises the primary forces in the undergraduate experience as located within the classroom setting. However, concentrating analysis on what happens in the classroom fails to account for a significant part of the undergraduate experience. As delivery modes expand and universities increasingly search for improved ways of providing a quality higher education experience, an instrument limited to classroom interactions is increasingly inadequate. If the standard instrument for course assessment is not able to truly measure the student experience across diverse settings, it may be a potential impediment to innovation and distinctiveness.

Two alternatives for modifying the CEQ provide a possible means of addressing this constraint. The first involves emphasising the broad quality assurance function of the existing instrument, but allows institutions to use or add their own more context-specific instruments to gather institutionally unique quality improvement information. Splitting the quality management process in this way may have a number of limitations. For instance, both the power and generality of the current instrument could be undermined and the lack of broader contextualisation of the institutionally-specific additional items may distort findings. Promoting the use of institutionally-based instruments also magnifies

the threat of relativism because innovative or distinctive approaches in teaching may not be readily measurable against conventional criteria. Further, it may encourage a limited understanding of the salient characteristics of the undergraduate experience and promote a narrowing of focus for policy and planning.

A second alternative for change guided the current extension process and involved extending the CEQ to address a more holistic range of factors in the contemporary undergraduate experience. This perspective recognised that students' learning experiences are embedded in a number of contextual situations that provide a 'dynamic web of influences', implying that the quality and outcomes of undergraduate study are composed of more than the quality of teaching alone (Pascarella, 1991, p. 458). It is supported by the assumption that a significant part of university experiences and learning outcomes are affected by what happens outside the classroom. A large body of research indicates that students' out-of-class experiences may have as much influence on the development of higher-order cognitive skills as do their more formal, classroom-based instructional experiences (Pascarella & Terenzini, 1991, 1998). Researchers such as Astin (1993), Pace (1979), Pascarella (1985), Terenzini and Wright (1987) and Tinto (1987) stress that engagement is fostered by a supportive learning environment. McInnis (1997) and McInnis and James (1995) also showed that these findings were relevant to the contemporary Australian context.

It became clear that the existing CEQ instrument needed to be extended to incorporate the social, technological, interpersonal and resource aspects of the university undergraduate experience. If this were to be achieved, then universities would have a broader range of items and sub-scales with which to reflect how input characteristics affect the undergraduate experience. With this motive, an exercise to extend the current CEQ was undertaken. Psychometrically modelled to resemble the original CEQ in a number of key respects, the new instrument was intended to connect current trends in educational theory with an understanding of the relevant aspects of current undergraduates' experiences, and to provide stable and reliable measurement of constructs not measured by the current instrument. A number of areas salient in the contemporary undergraduate experience not probed by the current CEQ were identified. These included:

   learning climate and intellectual environment;
   social dimensions of learning;
   provision and utilisation of resources to encourage and support independent learning;
   guidance and support for students to encourage increasing academic independence;
   analysis of higher-order graduate outcomes beyond generic skills;
   enhancement of graduates' intellectual stimulus and challenge;
   recognition of the growing importance of information technology;
   acknowledgement of the international context of learning.

Instrument Development Methodology

Extending the CEQ meant that a new version of the questionnaire was needed. Like its predecessor, the new instrument was required to be accurate, reliable and to have known tolerance or error limits. Wright and Masters (1982) defined the characteristics that measurement should have. They listed four requirements as direction, order, magnitude and replicable units. Each sub-scale must measure a specifically defined trait or domain of

interest. Just as the initial CEQ focused on specific areas of the learning experience of the undergraduate, the extended instrument would be required to measure additional domains or traits and do so with at least the same level of accuracy. Estimating these characteristics is called calibration. As Thurstone (1927) demanded, the trait being measured should not affect the measurement instrument and the measurement instrument should not affect the trait being measured. Furthermore, the measure obtained of the trait should not be affected by which instrument is used, given that we know the error and accuracy levels of the instrument. Any one of a set of equivalent instruments should give a measure of the trait consistent with measures obtained with any other equivalent instrument. If a scale is affected by the people who use it, or who are assessed by it, its validity is threatened. 'Within the range of objects for which the measuring instrument is intended, its function must be independent of the object of measurement' (Thorndike, 1971). This property is known as specific objectivity. For specific objectivity to hold, a measurement must be independent of the measuring instrument and the instrument must function independently of the traits measured.

Developing the Scales

Developing the items and the sub-scales involved several stages. The first stage involved a series of stakeholder meetings, student focus groups and literature reviews. In a second stage, outcomes of these processes were used to draft items and to propose how the item stems could be used to represent levels within the strands. Third, a national specialist group reviewed the item sets. This process, called panelling, was used to establish the face validity of item sets, eliminate linguistic ambiguities, and analyse the adequacy of the item pool. Fourth, the refined item sets were piloted at a small representative sample of universities. On-site data collection enabled the team to obtain feedback from more than 400 students on the items and directions for further improvements. To enable a pilot of the large number of draft items, four overlapping versions of the pilot instrument were developed, anchored by a series of common items including the original 25 CEQ items. Because of the assumed constant difference between the categories in a Likert scale, the Rasch (1960) rating scale model was used to examine the latent properties of the scales (Wright & Masters, 1982). The purpose of the analysis was to develop a series of unidimensional scales each consisting of items with invariant measurement properties. The analysis provided information on the estimates of item locations on the variable, of overlap and redundancy among items, and about items with over-deterministic or unstable measurement qualities. In combination with further substantive review by the national panel, deficient or misfitting items were refined or discarded. Final item and scale selections were then approved for a national trial in which a single set of items would be administered to all students sampled. Thirty items were developed in a series of six sub-scales, in addition to the existing CEQ (Table 1). The six final sub-scales were labelled 'student support', 'learning resources', 'course organisation', 'learning community', 'graduate qualities' and 'intellectual motivation'.

The final stage involved a nationally representative sample of students from all universities that agreed to participate. The sampling strategy involved stratifying the Australian university population by state, institution type and (ten) broad fields of study (DETYA, 1998). Guided by these specifications, 15 universities across all Australian states were included in the trial. A mixture of mail-out and on-site collection methods were used to collect a total of 3691 student responses. To counter and enable evaluation of response sets, order and fatigue effects, three versions of the questionnaire were employed, with each

TABLE 1. Item and Scale Descriptions and Labels

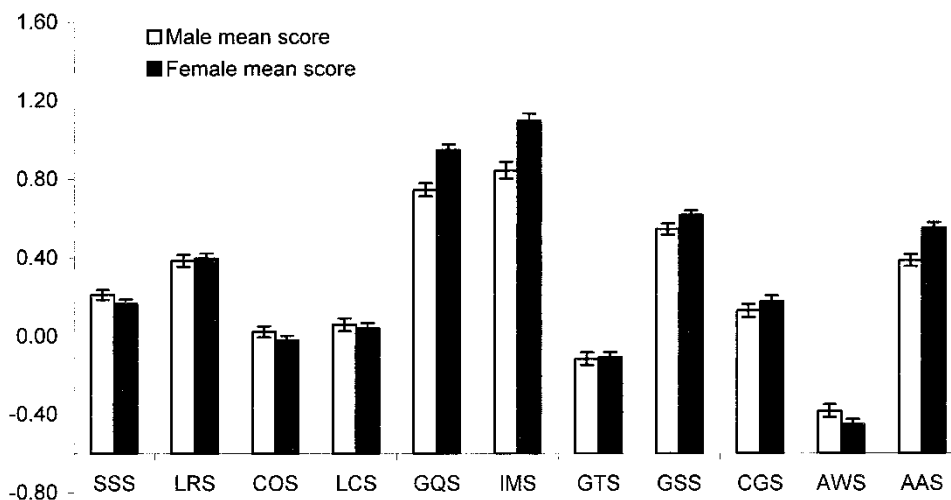|	Scales	Items
SSS	The Student Support Scale is concerned with access to, and satisfaction with, key university facilities and services supporting student learning outcomes.	76. The library services were readily accessible 77. I was able to access information technology resources when I needed them 79. I was satisfied with the course and careers advice provided 81. Health, welfare and counselling services met my requirements 39. Relevant learning resources were accessible when I needed them
LRS	The Learning Resources Scale consists of items primarily focussed on the appropriateness and effectiveness of sources of information and course materials.	75. The library resources were appropriate for my needs 78. Where it was used, the information technology in teaching and learning was effective 38. It was made clear what resources were available to help me learn 40. The study materials were clear and concise 71. Course materials were relevant and up to date
COS	The Course Organisation Scale contains items investigating the adequacy of the administrative structures and flexibilities of course organisation.	25. The course was well organised 26. I was given helpful advice when planning my academic program 27. The course content was organised in a systematic way 60. There was sufficient flexibility in my course to suit my needs 61. I had enough choices of the topics I wanted to study
LCS	The Learning Community Scale contains five items on student perceptions of the social experience of learning at university.	29. I felt part of a group of students and staff committed to learning 31. I was able to explore academic interests with staff and students 34. I learned to explore ideas confidently with other people 63. Students' ideas and suggestions were used during the course 30. I felt I belonged to the university community
GQS	The Graduate Qualities Scale has six items tapping qualities typically associated with higher order outcomes, especially attitudes and perspectives related to the relevance of the course for lifelong learning.	50. University stimulated my enthusiasm for further learning 68. The course provided me with a broad overview of my field of knowledge 51. My university experience encouraged me to value perspectives other than my own 54. I learned to apply principles from this course to new situations 55. The course developed my confidence to investigate new ideas 66. I consider what I learned valuable for my future
IMS	The Intellectual Motivation Scale is designed to identify perceptions of the impact of the course in inspiring and enabling individuals, as well as a global item enabling students to evaluate their overall university experience.	44. I found my studies intellectually stimulating 49. I found the course motivating 46. The course has stimulated my interest in the field of study 72. Overall, my university experience was worthwhile

FIG. 1. Mean scale scores with 95% confidence bands for men and women. [1]

form presenting the items in a different randomised sequence. The analyses were then able to check for the relationship between item sequence, model fit and missing data effects.

Measurement Properties of Items and Scales

The overall satisfaction measure was not necessarily the purpose of the scale so each of the six additional sub-scales was analysed separately. The Rasch reliability index was used to indicate the extent to which items were separated along each of the respective variables. The index has a range of 0.0 to 1.0 with a value of 1.0 indicating that each contributes uniquely to clarification of the variable. Given this, separation reliability can be understood as a measure of construct validity (Wright & Masters, 1982). The results indicated that all items had consistent measurement characteristics in relation to the latent traits, as their values ranged from 0.85 to 0.93.

A further series of analyses was also conducted to examine the behaviour of the new CEQ scales. The data were analysed concurrently with data from the existing CEQ scales and means for individual scales across sample subgroups were calculated. Given that, in a Rasch analysis, scale means are generally centred about zero, a positive score indicates an above average satisfaction of a group and likelihood that the group would on average choose positive categories. A negative score indicates the student group was slightly below the mean value and would, on average, choose less positive categories. Because the scales each measured different things, they may have naturally different levels of 'satisfaction demand' associated with them. While it was possible to compare differences between groups on the single scale or perhaps even trends across scales, it was not appropriate to compare differences between or within groups across different scales.

Scale Performances Across Student Groups

There was not a great deal of difference between gender groups on any of the scales, except the graduate qualities (GQS), appropriate assessment (AAS) and intellectual motivation (IMS) scales, as shown in Figure 1.

Similarities in scale trends across the fields of study were apparent, supporting the results reported by Pascarella and Terenzini (1991) and Davis and Murrell (1993). That is, areas such as veterinary science, in which formal teaching is highly rated, tend to be those in which learning was extended through social interaction beyond the classroom. While formal teaching was rated poorly in the field of engineering, for example, the learning community mean score indicated the presence of a learning environment extending beyond the classroom. A diagnostic interpretation of this result could lead to inquiry into these formerly undetected pedagogical modes.

It appears that the extent to which students feel intellectually motivated and stimulated by their university experience increased with age. However, the learning community scale data indicated a contrary trend. Older students felt less a part of a 'university community' that involved them 'working collaboratively with other students'. Students' impressions of generic skills obtained through university experience rise with the year of study. Mean scores on the graduate qualities scale also showed significant increases over year levels, indicating change in students' attitudes towards their courses over years. As students' expectations and application levels change, it may be that those things that appear to be impractical or even irrelevant in earlier years may emerge as more helpfully related to vocational and extra-curricular activities. That is they can 'apply principles from their course to new situations', feel 'stimulated by their course to engage with further learning' and consider 'what they learned as valuable for their future'. In contrast, the scores on the learning community scale decreased over years of study. It is possible that this may reflect the increasingly individual nature of academic involvements that follow higher education academic progressions.

Conclusion

Additional scales for the CEQ were developed and linked to the existing form of the instrument. The new scales broaden the instrument through attaching relevance to aspects of undergraduate student experience not measured by the existing instrument. The scales in the existing instrument were supplemented by scales such as 'student support', 'learning resources', 'learning community', 'graduate qualities', and 'intellectual motivation'. These new scales were shown to be psychometrically reliable and accurate. No item was shown to misfit under a Rasch model analysis and the scales show reasonable discrimination across fields of study. Analyses also show that the outcomes are consistent with other studies of the undergraduate experience and supportive of the results of the previous CEQ instrument and sub-scales. Moreover, the extension of the CEQ now incorporates several extensions of the classroom emphasis of the existing instruments and allows fields of study between universities to make comparisons of student satisfaction in a total of 11 areas if all sub-scales are used. The new scales both reinforce and embellish results produced by the current CEQ. They provide a mechanism for measuring a broader range of learning experiences in a university context.

In addition to analysing the extent to which students feel positive about structured teaching and assessment methods, the new scales provide a mechanism for measuring more general pedagogical contexts and higher-order outcomes of undergraduate programmes. While support for a scale measuring overall student satisfaction was not obtained, Rasch calibration indicated that all of the six scales produced have precise and stable measurement properties. In these examinations, the new scales were found to both reinforce and enhance results produced by the current CEQ scales.

Note

[1] GTS–"Good Teaching"; GSS–"Generic Skills"; CGS–"Clear Goals"; AWS–"Appropriate Workload".

References

ASHENDEN, D., & MILLIGAN, S., 2000, *The Good Universities Guide* (Subiaco, WA, Hobsons).

ASTIN, A., 1993, *What Matters in College? Four critical years revisited* (San Francisco, CA, Jossey-Bass).

DAUS, T.M. AND HURRELL, P.H., 1993, 'A structured model of perceived academic, personal, and vocational gains related to college student responsibility', *Research in Higher Education, 34(4)*, 267–89.

DETYA, 1998, *Selected Higher Education Student Statistics* (Canberra, Australian Government Publishing Service).

MARSH, H.W., 1987, 'Students' evaluations of university teaching: research findings, methodological issues, and directions for future research', *International Journal of Educational Research, 11*(3), pp. 253–288.

MCINNIS, C., 1997, 'Defining and assessing the student experience in the quality management process', *Tertiary Education and Management, 3*, 63–71.

MCINNIS, C. & JAMES, R., 1995, *First Year on Campus* (Canberra, Australian Government Publishing Service).

PACE, C., 1979, *Measuring Outcomes of College: Fifty years of findings and recommendations for the future* (San Francisco, CA, Jossey-Bass).

PASCARELLA, E., 1985, 'College environmental influences on learning and cognitive development: a critical review and synthesis' in Smart, J., (Ed.), *Higher Education: Handbook of theory and research, Vol. 1.* (New York, Agathon).

PASCARELLA, E., 1991, 'The impact of college on students: the nature of the evidence', *The Review of Higher Education, 14*(4), 453–66.

PASCARELLA, E., & TERENZINI, P., 1991, *How College Affects Students* (San Francisco, CA, Jossey-Bass).

PASCARELLA, E., & TERENZINI, P., 1998, 'Studying college students in the 21st century: meeting new challenges', Review *of Higher Education, 21*, 151–65.

RAMSDEN, P., 1991a, 'A performance indicator of teaching quality in higher education: the Course Experience Questionnaire', *Studies in Higher Education, 16*(2), 129–49.

RAMSDEN, P., 1991b, 'Report on the CEQ Trial', in Linke, R. (Ed.), *Performance Indicators in Higher Education, Vol. 2.* (Canberra, Australian Government Publishing Service).

RASCH, G., 1960, *Probabilistic Models for Some Intelligence and Attainment Tests.* (Chicago, IL, MESA Press).

TERENZINI, P., & WRIGHT, T., 1987, 'Students' personal growth during the first two years of college', *Review of Higher Education, 10*(2), 259–71.

THORNDIKE, R.L., 1971, *Educational Measurement* (2nd edition.). (Washington, DC, American Council on Education).

THURSTONE, L.L., 1927, 'Law of comparative judgement', *Psychological Review, 34,* 278–86.

TINTO, V., 1987, *Leaving College: Rethinking the causes and cures of student attrition.* (Chicago, IL, University of Chicago Press).

WRIGHT, B. D., & MASTERS, G. N., 1982, *Rating Scale Analysis.* (Chicago, IL, MESA Press).