

Using the Course Experience Questionnaire for evaluating undergraduate tourism management courses in Greece

Dimitrios P. Stergiou

School of Social Sciences, Hellenic Open University, Karatza 88, Patras 26226, Greece, tel: 00306942050223, email: dstergiou@eap.gr

Dr. Dimitrios P. Stergiou is an Associated Academic at the Hellenic Open University, Greece. He obtained his PhD in tourism education at the University of Surrey in 2004. His research and consultation works are in the fields of tourism and hospitality management and tourism education.

David Airey

School of Management, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom, tel: +44 (0)1483 689656, email: d.airey@surrey.ac.uk

David Airey is Professor of Tourism Management at the University of Surrey where, until he retired from his full-time position in 2009, he also served as Pro-Vice Chancellor. His main areas of academic interest are in tourism education and tourism policy and organisation. He currently co-chairs the UN World Tourism Organization's Education and Science Council and in 2006 received the UNWTO's prestigious Ulysses award for his contribution to tourism education.

Using the Course Experience Questionnaire for evaluating undergraduate tourism management courses in Greece

ABSTRACT

An adapted version of the Course Experience Questionnaire (CEQ) was administered to tourism management students at two technological educational institutes in Greece. This questionnaire has been previously utilized with students taking tourism-related degree courses in the United Kingdom. The analyses presented herein focus on the psychometric properties of the adopted research instrument with the purpose of testing its applicability in the context of tourism higher education in Greece. The results suggest that the CEQ demonstrated an appropriate five-factor structure, satisfactory internal consistency and appropriate relationships with students' rating of their overall satisfaction. Limitations of and directions for future research are discussed.

Keywords: student experience, course evaluation, tourism education

INTRODUCTION

Systems for the evaluation of teaching and course quality in higher education institutions (HEIs) have long been established both in the United States (US) and Australia and they have also become increasingly common in the United Kingdom (UK). In a recent comparative review of course evaluation surveys in these countries, the Higher Education Academy (Hanbury, 2007, p. 1) stated that the importance of gaining systematic evaluations of courses by means of national surveys “reflects the growing focus on the quality of teaching, formal institutional arrangements, and a growing emphasis on competition between higher education institutions”. Accordingly, within Greece, there has been an upsurge of interest in this area from a range of different perspectives driven both internally by institutions themselves and externally by national quality initiatives and general public calls for increased accountability and quality assurance.

Whilst there is a large number of possible sources of evaluation data on both teaching and course quality, the commonest form of input to educational evaluation (particularly in the US, the UK and Australia) is feedback from students (Hoyt and Perera, 2000). Indeed, the practice of obtaining student feedback on individual teachers and course units is widespread and causes little concern in these countries although the uses to which it is put may be controversial. For example, the Course Experience Questionnaire (CEQ) is an annual government-mandated survey which all Australian Universities take part in, and is sent to the previous year’s graduating students. In the UK, the

National Student Survey (NSS) is administered to students still in their final year of undergraduate studies. The findings from both surveys are used to identify problem areas and inform enhancement activities.

In the Greek context, Law 3374 of 2005 provides the legislative framework for the evaluation of HEIs. However, systematic collection and processing of evaluative data is not well established in most Greek universities. Individual lecturers may voluntarily obtain feedback on their teaching from students using questionnaires of various kinds. At the institutional level, some feedback is gathered for the purposes of departmental evaluations but often on an incomplete and *ad-hoc* basis. Indeed, four years following the passing of Law 3374, only 5 out of 500 departments of higher education institutions have completed evaluation processes (Kalimeri, 2009). It can be argued, therefore, that there is a clear need for a more standardised approach to this process; all the more so as course evaluation is promoted by the Greek government as a decisive step towards the convergence of the HE system with the principles that govern the European Higher Education Area.

Against this background, the Greek Ministry of Tourism Development, sharing responsibility for monitoring and evaluation of tourism educational programmes with the Ministry of Education, set up a project to advise on the development of such a survey. This led in turn to the commissioning of a pilot study of current undergraduate students following tourism management programmes of study. This was carried out during the spring of 2008 using an adapted version of the

CEQ and yielded responses from 283 students at two institutions. In line with the use and purpose of the CEQ (see McInnis, Griffin, James and Coates, 2001), the objective of this project was to provide reliable and valid indicators of crucial aspects of students' course experiences and ratings of overall satisfaction. An additional objective was to determine whether the questionnaire used in this study could be standardised as a performance indicator for monitoring the quality of tourism management academic programmes in Greece. This paper will focus on this second objective and emphasis is placed on analyses concerning the psychometric properties of the adopted research instrument. These analyses are important for HEIs to have confidence in the use of the questionnaire.

OBTAINING STUDENT FEEDBACK IN HIGHER EDUCATION

Student evaluation in higher education can take place at various levels: at the level of individual teachers, course units, programmes of study, departments and institutions. As Richardson (2005, p. 402) has pointed out, "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". Clearly, the level at which one should collect feedback is dependent upon the purpose of the investigation. From the perspective of this study, the focus is on the experience of students over a whole programme of study, rather than an individual module. Accordingly, this section draws primarily on the predominantly Australian and British literature that is concerned with students' evaluations of their programmes.

Although formal questionnaires are most often used to obtain student feedback in higher education, they by no means constitute the only method. Student feedback can also be collected in many other ways. These include informal class discussions, formal staff-student meetings, students' notes, diaries and/or log books, student interviews and focus groups (Keane and Mac Labhrainn, 2005). However, formal questionnaire surveys have two methodological benefits: they can provide an opportunity to collect data from the entire student population; and they document the experiences of the student population in a systematic way, providing useful time-series data (Richardson, 2005). These two benefits are afforded by such instruments being standardised, with associated psychometric properties of reliability and validity, thus aiming at collecting more quantitative-type data (Hanbury, 2007). Student feedback could, of course, be obtained by means of open-ended questionnaires. Nevertheless, whilst rich and informative, the analysis of open-ended responses and other qualitative data may prove an extremely time-consuming and labour-intensive effort and is, therefore, not used for course monitoring when surveying large numbers of students (Keane and Mac Labhrainn, 2005).

Much of the research evidence in this area has been concerned with the reliability and validity of students' evaluations of their programmes (Prebble et al., 2004). To this extent, formal student surveys typically contain groupings of items reflecting different dimensions of the student experience of a particular course, referred to as scales. Reliability and validity are important psychometric properties of surveys, with reliability being concerned with the accuracy of the

actual measuring instrument, and validity referring to the instrument's success at measuring what it purports to measure (Nunnally, 1978; Hinkin, 1995). Tests of reliability and validity are performed on the scales, and as such assess the magnitude of measurement errors in survey data (Bound, Brown and Mathiowetz, 2001). A statistical technique called factor analysis is also typically used to establish whether the groups of items form the latent structure (dimensions) they were supposed to. Tests of reliability and validity are, of course, specific to the survey and as such will be discussed below with regard to the survey presented in this study. However, overall, Marsh (1987), Paulsen (2002) and Richardson (2005) suggest that student ratings demonstrate acceptable psychometric properties, and can provide important evidence for educational research.

The instrument that has been most widely used in published work is Ramsden's (1991) Course Experience Questionnaire (CEQ). In completing this questionnaire, students are required to note the extent of their agreement or disagreement with a set of 31 items on a five-point Likert scale, ranging from 'definitely agree', scoring five, to 'definitely disagree', scoring one. Statements 1-30 are intended to reflect five aspects of perceived teaching quality on particular academic programmes: Good Teaching, Clear Goals and Standards, Appropriate Workload, Appropriate Assessment, and Emphasis on Student Independence. Statement 31 measures the respondents' overall level of satisfaction with their programmes.

The CEQ has a substantial literature addressing its reliability and validity (e.g. Ramsden, 1991; Richardson, 1994; Richardson, 2005; Hanbury, 2007). In the Australian national trial of the CEQ (Ramsden, 1991), its scales have been found to have generally satisfactory reliability levels with Cronbach alpha coefficients ranging between 0.71 for the Appropriate Assessment scale, to 0.87 for the Good Teaching scale. The scales of the CEQ have also been tested with British students and were found to have slightly lower Cronbach alpha coefficients (Richardson, 1994). However, as noted by Richardson (2005), there is no evidence on their test-retest reliability. The construct validity of the CEQ according to the results of factor analyses is also broadly satisfactory, with most items loading on distinct factors reflecting their assigned scales (Hanbury, 2007). The fact that the modal solution is a single factor on which all scales show significant loadings also constitutes evidence for CEQ's construct validity (Richardson, 2005). The criterion validity of the CEQ is supported by statistically significant correlations between the scales and ratings of overall satisfaction (Byrne and Flood, 2003). Finally, the discriminant validity of the CEQ is illustrated by the fact that respondents' scores on the five scales vary across different broad fields of study or across different institutions offering programmes in the same field (Richardson, 2005).

Möller (2002; also see Downie and Möller, 2002) adapted the CEQ for use with students who were taking degree courses in hospitality, leisure, sport and tourism subject areas in the UK. This adapted version was administered to 634 campus-based students at six HEIs during the summer of 2001, to identify

stronger and weaker areas of teaching on degree programmes within these subject areas. The original version of the CEQ was modified in certain respects. Firstly, a Generic Skills scale was added to investigate the skills that students gained from their programmes. This scale was used in earlier versions of the CEQ, administered by the Graduate Careers Council of Australia (GCCA), in response to concerns about the employability of graduates (Ainley and Long, 1994). Secondly, the Emphasis on Independence Scale was dropped. Möller does not provide any explanation for this decision; however the scale was probably removed on the grounds that previous factor-analytic research has indicated a consistent tendency for items on this scale to load on other factors (see Richardson, 2005). And, thirdly, the Good Teaching scale was split into two groups, Teaching, and Academic Environment, to “differentiate between the more direct and the more general aspects of teaching” (Möller, 2002, p. 1). The six scales used were: Academic Environment, Teaching, Skills Development, Appropriate Assessment, Appropriate Workload, and Clear Goals and Standards. Following this pilot study, slightly modified versions of the questionnaire were used to conduct nationwide student surveys and main findings from these have been published for the years 2002-2005 (Hospitality, Leisure, Sport and Tourism Network, 2009). Since no reliability or validity for these instruments was reported, the emphasis here is on the 2001 pilot survey, which formed the basis for future surveys.

As in previous research, responses assigned to the relevant items of the CEQ by each respondent were summed to obtain a score on each subscale. Analysis of the mean scores and standard deviations for these 634 students on the subscales of the CEQ, suggested that the six groups of teaching were ranked in the same order by all four subject areas and that all scales got very similar scores. Skills Development consistently got the highest rating, followed by Academic Environment, Teaching, Appropriate Assessment, Clear Goals and Standards, and Appropriate Workload. Möller also carried out statistical tests which compared the sets of data obtained by the participating students. Using 5% as the criterion of statistical significance, Möller found significant overall differences between Skills Development and Appropriate Workload/Clear Goals and Standards. Based on these results Möller (2002, p. 13) concluded that the adapted version of the CEQ provides “a good means of determining strengths and weaknesses of teaching on degree programmes in our subject areas”. The generality of this conclusion, however, may be limited, as no sample outside the UK has been studied. In addition, Möller presented no evidence to support the reliability and validity of this instrument.

In Greece, recent discussions about national educational evaluation measures, including measures for tourism studies, have lead to an upsurge of interest in measures of teaching and course quality. In fact, in 2007 the Greek Ministry of Economics and Finance (p. 56) referred to the need for higher education institutions to develop “standards of quality assurance and methodologically adapt them to the particularities of their taught subjects”. In this connection,

Möller's previous experience with the CEQ adapted for use with students in tourism-related courses provides a useful starting point for the purposes of the present investigation. More specifically, the analyses herein focus on the factor structure of Möller's (2002) version of the CEQ for a new sample of undergraduate tourism students, drawn from two HEIs in Greece. Follow-up analyses test the internal consistency reliabilities of the revised scales. Regression analyses are then employed to test the relationship of these scales to overall satisfaction with course. In the light of the changes made to the CEQ by Möller, it was necessary to evaluate its psychometric properties and to establish its constituent structure in this distinctive situation.

METHOD

Context

After about 35 years of development, Greece now has a fairly well-developed higher education system in tourism which, in common with Western Europe, has experienced significant expansion in the past few years. According to Greek legislation, higher education consists of two parallel sectors (Ministry of National Education and Religious Affairs, 2005):

- The university sector, which includes universities, polytechnics, and the Athens School of Fine Arts. Within this sector, degree-level courses with a management component relating to tourism (but not degrees in tourism) are offered by the Business Administration Departments of the University of the Aegean and the University of Patras. Also in the

university sector, three universities (University of the Aegean, University of Piraeus, and the Hellenic Open University) offer postgraduate programmes in tourism, leading to Master's or PhD degrees. These have been established during the past ten years.

- The Technical Sector, which includes the Technological Educational Institutions (TEIs). There are eight TEIs (in Athens, Epirus, Heraclion, Lamia, Larissa, Patras, Piraeus, and Thessalonica) offering vocational courses leading to an undergraduate degree in tourism management. It needs to be emphasised that TEIs are very similar to the former British polytechnics (Christou, 1999). They were fully integrated into the higher education system in 2001.

The provision in the two sectors noted here provides a fairly comprehensive system of tourism higher education and the recent introduction of postgraduate degrees indicates the way in which the system is expanding to meet the needs of students and of the tourism industry.

Samples

Data for this survey were collected from students registered in two tourism management courses. These were located in the tourism management departments of two TEIs, situated in different geographical departments of Greece, which suited the convenience of the researchers by virtue of their accessibility. It should be noted that all students admitted to TEIs should have successfully taken the national examinations organised by the Ministry of

National Education and Religious Affairs or the special examinations organized by TEIs. Moreover, similar departments have very similar admission criteria and offer broadly the same curricula. This results in the admission of similar students and the offering of similar courses of study. In this connection, the TEIs selected for this study are comparable to others, so representativeness was not an issue in terms of selecting which ones should participate in the study. For reasons of institutional confidentiality, the TEIs will not otherwise be identified in this study. It should also be pointed out that during the time at which the present research was conducted, TEIs acted as the sole provider of tourism degrees at undergraduate level. Data collection took place during the spring of 2008.

No reliable data exists on the total size or characteristics of the student population currently studying modules in the undergraduate tourism management programmes of the participating TEIs and this posed problems for data collection. Following this, a decision was reached to collect data by asking students to respond to the CEQ during normal classroom periods at their desks under the supervision of two researchers. Data were available from 283 students who had been invited to contribute to a study of 'tourism course evaluation'. Within this sample, 59.3% of the students were female and 41.7% were male. Their ages ranged from 21 years to 32 years, with an overall mean of 23.3 years. There were no significant differences between the two sub-samples on any of the demographic variables collected for this study.

Therefore, all presented analyses were based on a total sample of 283 respondents.

Measure

The CEQ used in the current study was based on the 31-item instrument used by Möller (2002). The introduction to the CEQ explained the aim of the survey and assured participants that their responses would be kept confidential and anonymous. Respondents were asked to think about their course as a whole rather than about individual units, topics, or lecturers. For each item, the participants were asked to indicate their level of agreement or disagreement with the relevant statement using a five-point scale from 5 for 'definitely agree' to 1 for 'definitely disagree'. Statements 1-30 fell under the six scales identified by Möller as reflecting different dimensions of teaching (see Appendix A). Half of these items referred to positive aspects, whereas the other half referred to negative ones and were scored in reverse. Statement 31 was concerned with students' overall satisfaction.

RESULTS

Confirmatory factor analysis

Although exploratory factor analysis (EFA) provides a powerful tool for assessing the extent to which a set of items assesses a particular set of scales, a major weakness of this technique is the inability to quantify the goodness-of-fit of the resulting factor structure (Long, 1983). In addition, EFA involves a post hoc interpretation of the results, whereas confirmatory factor analysis (CFA)

specifies a priori relationships among the scales or variables of interest (Hinkin Tracey and Enz, 1997). In the case of the CEQ prior analyses were already available in the literature. Consequently, CFA was chosen over EFA.

Confirmatory factor analysis is a type of structural equations analysis that is designed to assess the goodness-of-fit of rival models (Kim and Mueller, 1978). Richardson (2005) provided an exhaustive review of CEQ studies using exploratory and confirmatory factor analytic methods. He concluded that responses to the CEQ can be conceptualised at different levels. At the most general level he described a single common factor model of students' perceptions of academic quality. In addition, he reported multi-trait models with the number of factors equal to the predicted number of CEQ constructs. Moreover, the use of negatively worded (reverse-scored) items in the scale presented an issue of concern. Reverse-scoring of items "has been shown to reduce the validity of questionnaire responses and may introduce systematic error to a scale" (Hinkin, 1995, p. 972). More specifically, researchers have shown that they may result in two response factors, one related to positively worded items and the other related to negatively worded items (e.g. Schmitt and Stults, 1985; Woods, 2006). Following these, the purpose of the analysis was to assess the goodness-of-fit of three models: a single common factor model, a two-factor model, and the hypothesized six-scale factor structure.

As there seems to be little consensus on what are the appropriate indices to assess confirmatory factor analytic results, several widely accepted goodness-

of-fit indices were computed. While frequently reported, the chi-square statistic was not used in the evaluation of model fit as it is known to be quite sensitive to sample size (Hinkin et al., 1997). The main criteria used to judge model fit included Bentler's (1990) comparative fit index (CFI) and the Tucker-Lewis Index (TLI; Tucker and Lewis, 1973), for which values greater than 0.85 indicate reasonably good model fit (Widaman, 1985); and the root mean square error of approximation (RMSEA), with a value of less than 0.08 considered acceptable (Spector, 2001). Ginns (2003) had earlier, and for similar reasons, recommended use of these indices in relation to the scale structure of an adapted version of the CEQ.

Following Harvey, Billings and Nilan's (1985) recommendation, the confirmatory factor analyses of the 30 items were conducted by using the item variance-covariance matrix. The fit indices for the three models are presented in Table 1. Scrutiny of the fit indices indicates a poor fit overall of the one-factor model to the collected data. The TLI, for example, is 0.744. It was then tested whether the two-factor model represented a better fit to the data than the one-factor model. The goodness-of-fit indices showed that the two-factor model fits the data relatively well. The TLI of 0.893, for example, shows that the remaining improvement in fit possible (0.107) for the two-factor model is not impressive. Using maximum likelihood (ML) estimates, a two-factor solution ensued with the first factor comprising the positively worded items and the second factor comprising the negatively worded items.

[Table 1 about here]

The fit of the six-factor model was evaluated using the sample variance-covariance matrix and a ML solution. Table 1 presents the three fit indices (described earlier) of the measurement model. The CFI was 0.811, the TLI was 0.787 and the RMSEA was 0.106. As these indices were not within the range of conventionally accepted values, the six-factor model was not supported. However, Hinkin et al. (1997) point out that if the model does not fit well, modification indices for the lambda matrix (i.e. a matrix that shows how well the observed variables 'load on' the latent variables or scales) may be used to guide respecification (i.e. how to improve the model). For CFA, one sensible solution is to eliminate items that load on multiple factors. According to Medsker, Williams and Holohan (1994), values less than four are acceptable for defining a factor, whereas values higher than five indicate that the items are loading on multiple factors. The modification indices showed that 11 items exceeded the suggested cut-off value. Thus, all six Academic Environment items were eliminated, two items for each of the Teaching and Skills Development scales were eliminated, and one item was eliminated for each of the Appropriate Assessment and Clear Goals and Standards scales. The remaining five factors were defined by 19 items: four Teaching items; five Skills Development items; four Appropriate Workload items; three Appropriate Assessment items; and three Clear Goals and Standards items (see Appendix A). Results from a CFA of the revised scales supported a five-factor model. Using the sample variance-covariance matrix and a ML solution, the CFI was

0.954, the TLI was 0.943 and the RMSEA was 0.042. These values suggest good fit for the five-factor model. In addition all modification indices were low. These results do not confirm that the items used by Möller (2002) constitute the six scales as intended.

Reliability assessment

After the CFA has been conducted and all items that were not consistent with the corresponding construct domain have been deleted, the reliabilities for the revised measure and scales should be calculated (Hinkin et al., 1997). Reliability is essentially a synonym for consistency (Cohen, Manion and Morrison, 2000). The researcher needs to be sure that the measuring instrument will behave in a fashion that is consistent with itself; that a very high proportion of the score on every occasion is due to the underlying scale variable, with a minimum of error (Oppenheim, 1992).

Although reliability may be measured in a number of ways, the most commonly accepted measure is internal consistency reliability using Cronbach's (1951) alpha coefficient. Following classical scaling theory, a scale will be internally consistent if the items correlate highly with each other – in which case they are also more likely to measure the same homogenous variable (Oppenheim, 1992). Items that are reliable, that is items with low error components, are more likely to fulfill these requirements. Since Cronbach's alpha measure provides an estimation of the proportion of the total variance that is not due to error, this represents the reliability of the scale. A large coefficient alpha ($\alpha > 0.70$)

provides an indication of strong item homogeneity and suggests that the sampling domain has been adequately captured (Nunnally, 1978; Churchill, 1979).

The calculation of the internal consistency reliabilities for the 19 retained CEQ items provides an indication of strong item homogeneity (see Table 2). The total scale reliability was a large coefficient alpha of 0.88, which is considered very good in educational research. As the CEQ contains nominated dimensions, it was also necessary to check the internal consistency of each dimension. This analysis resulted in respectable alpha coefficients (0.82 to 0.88) for all revised scales.

[Table 2 about here]

Criterion validity

The strong internal consistency reliability for the revised scales indicates that the retained items measure the same constructs, thus providing supportive evidence of construct validity. Further evidence of construct validity can be accomplished by demonstrating the existence of relationships with variables that are hypothesized to be outcomes of the focal measure (criterion validity; Cronbach and Meehl, 1955). Hypothesized relationships are usually confirmed using either correlation or regression analysis (Hinkin, 1995).

For the criterion validity assessment, item 31 ('Overall, I am satisfied with the quality of this degree course') was included in the CEQ to validate its use as a measure of perceived quality. The response choices ranged from 1 (strongly disagree) to 5 (strongly agree). The overall mean response was 3.04 and 56% of the respondents indicated their agreement with the statement, both findings implying a moderate degree of satisfaction with TEI courses. The correlation coefficients between the revised scales and student responses to item 31 were: Teaching, +0.67; Skills Development, +0.73; Appropriate Assessment, +0.34; Appropriate Workload, +0.30; and Clear Goals and Standards, +0.43, $p < 0.01$ in each case. In other words, students' level of general satisfaction with TEI courses was determined rather more by their perceptions of skills development and teaching than by their perceptions of clear goals and standards, appropriate assessment, and receiving an appropriate workload.

To assess the criterion validity of the revised scales, a multiple regression analysis was also undertaken. More specifically, Johnson's (2000) method for estimating the relative importance of predictor variables in multiple regression was utilised. Although standardised regression coefficients and squared beta weights are often used in this respect, many authors have maintained that these indices are inadequate in this case because they fail to consider both the effect the variable has by itself and in combination with the other variables in the model (Budescu, 1993; Johnson, 2001). Instead, the measure adopted here has been specifically designed to be used when "the researcher is interested in the relative contribution each variable makes to the prediction of a dependent

variable, considering both its unique contribution and its contribution when combined with other variables” (Johnson, 2000, p. 2). Moreover, this method yields similar results to other regression methods with less computation (LeBreton, Binning, Adorno and Melcher, 2004).

The results of the Johnson analytic method are expressed as the percentage of variance each predictor accounts for (R^2) in the criterion out of the total percentage of variance accounted for by the entire set of predictors. Table 3 gives the percentage of R^2 explained in Overall Satisfaction (Item 31) by each of the revised scales. Following these results, Skills Development and Teaching emerged as the best predictors of Overall Satisfaction, followed by Appropriate Assessment, Clear Goals and Standards and Appropriate Workload. This is obviously a very similar pattern to that obtained in the correlation analysis, and this provides further support for the construct validity of the revised CEQ scales.

[Table 3 about here]

CONCLUSIONS

This appears to constitute the first investigation to provide evidence concerning the psychometric properties of the CEQ, using data from students in tourism programmes of study. Indeed, the most important finding from a methodological point of view is that the CEQ appears to be robust when it is administered in the context of tourism higher education in Greece. Used at the

TEIs, it exhibited an appropriate five-factor structure, satisfactory internal consistency and appropriate relationships with students' rating of their overall satisfaction.

The current study demonstrated support for five of the six CEQ scales proposed by Möller (2002). However, a closer inspection of the items revealed a narrower behavioural operationalisation of Teaching, suggesting that this scale has been too broadly defined as shown in Appendix A. This dimension was represented by four items. These items demonstrate individualised consideration behaviours such as “the teaching staff normally give me helpful feedback on how I am doing” and “the teaching staff of this course motivate me to do my best work”. Two other items, “my lecturers are extremely good at explaining things” and “the teaching staff work hard to make their subjects interesting”, were not judged to be empirically consistent in the CFA and appear to be conceptually inconsistent with the individualised consideration theme associated with the four retained items.

Equally important, from a methodological perspective, is the lack of support for the Academic Environment scale. There is an important lesson here, as suggested by DeVellis (2003), that researchers may underestimate the importance of reliability and validity to sound measurement, and may rely on face validity if a measure appears to be a valid index of a content domain. Clearly, the Academic Environment scale was not the result of sound scale development procedures. As Hinkin (1995, p. 982) stated, “given the desire to

complete research for submission for publication, the development of sound measures may not seem like an efficient use of a researcher's time". It has been shown here, however, that a measure may appear to capture a domain of interest, but lack construct validity.

In suggesting that the five-factor model received a good deal of support from the presented data, there is no intention to imply that in all cases and at all times this model best represents the evaluative judgements of tourism students. Indeed, there are several questions about these issues that are worthy of further examination. For one, a change in the educational context might change the pattern of support for one or more of these models and scales. For example, the achievement of the highest alpha value in this study for the Skills Development scale is undoubtedly associated with the vocational nature of the participating courses. If one studied courses embracing a more open agenda than the vocational courses examined here, different results could be found. Another issue of concern regards the wording of the items. In the two-factor model, all positively worded items formed one factor and all negatively worded items formed the other factor. To avoid or to establish the criticism that the two-factor structure is simply a function of item wording further research is needed. A more general validity issue concerns the modal (one-factor) solution. Students' evaluations of their courses have been found to be multidimensional. However, they were not dominated by a single overarching factor as has been the case in previous studies (see Richardson, 2005). In the context of this research it is therefore not a straightforward matter to plausibly interpret the

adopted version of the CEQ as an indicator of perceived academic quality. Thus, before dismissing any of these models further research is necessary.

Even more important, however, is the need is to expand the scope of research about educational evaluation issues. The present research represents an important advance in that it competitively tests various CEQ models in the context of tourism education. However, even though various models have been tested, questions having to do with why CEQ scales impact satisfaction outcomes in the way that they do have not been addressed here. For example, it was found that the five-factor model provided the best fit to the collected data. But why is it that Appropriate Assessment seems to be a more important predictor of overall satisfaction than Appropriate Workload? In any case, future research needs to move beyond “demonstration” and address the reasons why specific components of course evaluation affect different outcomes.

From the practical perspective of evaluating tourism courses, the confirmation of the five CEQ factors as reliable and valid measures of perceived academic quality in the tourism course context is an important finding. In the current study the CEQ was well accepted and understood by tourism students and may, therefore, be useful for evaluating the strengths and weaknesses of tourism courses and for national projects which focus on the five main factors which have now been replicated. From this can follow more in-depth investigations of the nature and causes of particular strengths and weaknesses following other methodological approaches. It is hoped that the current work is useful for

facilitating future development of the measurement of tourism course evaluation.

REFERENCES

Ainley, J., & Long, M. (1994). *The course experience survey 1992 graduates*.

Canberra: Australian Government Publishing Service.

Bentler, P.M. (1990). Comparative fit indexes in structural models.

Psychological Bulletin, 107, 238–246. doi:[10.1037/0033-2909.107.2.238](https://doi.org/10.1037/0033-2909.107.2.238)

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J.J. Heckman, & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5) (pp. 3705-3843). Amsterdam: Elsevier.

Budescu, D.V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542-551. doi:[10.1037/0033-2909.114.3.542](https://doi.org/10.1037/0033-2909.114.3.542)

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, 25, 221-236.

Christou, E. (1999). Hospitality management education in Greece: An exploratory study. *Tourism Management*, 20, 683-691. doi:[10.1016/S0261-5177\(99\)00039-4](https://doi.org/10.1016/S0261-5177(99)00039-4)

Churchill, G.A., Jr. (1979). A paradigm for developing better measures of

marketing constructs. *Journal of Marketing Research*, 16, 64–73.

[doi:10.2307/3150876](https://doi.org/10.2307/3150876)

Cohen, L., Manion, L., & Morrison, K. (2001). *Research methods in education* (5th ed.). London: Routledge-Falmer.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests.

Psychometrika, 16, 297-334. [doi:10.1007/BF02310555](https://doi.org/10.1007/BF02310555)

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests.

Psychological Bulletin, 52, 281-302. [doi:10.1037/h0040957](https://doi.org/10.1037/h0040957)

DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed.).

Thousand Oaks: Sage.

Downie, N., & Möller, I. (2002). The Ramsden Course Experience

Questionnaire: A pilot study of final-year students taking hospitality, leisure, sport and tourism degree courses. *Journal of Hospitality, Leisure, Sport and*

Tourism Education, 1(1), 77-81. [doi:10.3794/johlste.11.10](https://doi.org/10.3794/johlste.11.10)

Ginns, P. (2003). Scale structure of the student course experience

questionnaire over field and level of study. Paper presented at the Higher Education Research and Development Society of Australasia (HERDSA)

Conference, July 6-9, Christchurch: New Zealand.

Harvey, R.J., Billings, R.S., & Nilan, K.J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology*, 70, 461-468. doi:10.1037/0021-9010.70.3.461

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

Hinkin, T.R. (1995). A review of scale development practices in the study of organisations. *Journal of Management*, 21(5), 967-988.
doi:10.1177/014920639502100509

Hinkin, T.R., Tracey, J.B., & Enz, C.A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research*, 21(1), 100-121.

Hospitality, Leisure, Sport and Tourism Network (2009). Student course experience survey. Retrieved February 17, 2010, from
http://www.heacademy.ac.uk/hlst/projects/detail/ourwork/OP_student_course_experience_survey

Hoyt, D., & Perera, S. (2000). *Validity of the Idea student ratings of instruction system: An update, Idea research report #2*. Kansas: Kansas State University.

Johnson, J.W. (2000). A heuristic method for estimating for estimating the relative weight of predictors in multiple regression. *Multivariate Behavioural Research*, 35, 1-19. doi:10.1207/S15327906MBR3501_1

Johnson, J.W. (2001). RWEIGHT: Comparing the relative weight of predictors in a regression equation. *Applied Psychological Measurement*, 25(4), 342. doi:10.1177/01466210122032190

Kalimeri, H. (2009). Delays in the evaluation of TEIs and AEIs [in Greek]. Retrieved March 22, 2010, from <http://www.imerisia.gr/article.asp?catid=12333&subid=2&pubid=11591180>

Keane, E., & Mac Labhrainn, I. (2005). Obtaining student feedback on teaching and course quality, Centre for Excellence in Learning and Teaching, Briefing paper 2. Retrieved February 12, 2010, from http://www.nuigalway.ie/celt/documents/evaluation_ofteaching.pdf

Kim, J.O., & Mueller, C.W. (1978). *Factor analysis: statistical methods and practical issues*. London: Sage.

LeBreton, J.M., Binning, J.F., Adorno, A.J., & Melcher, K.M. (2004). Importance of personality and job-specific affect for predicting job attitudes and withdrawal behavior. *Organisational Research Methods*, 7(3), 300-325. doi:10.1177/1094428104266015

Long, J.S. (1983). *Confirmatory factor analysis*. Beverly Hills: Sage.

Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388. doi:[10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)

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

Medsker, G.J., Williams, L.J., & Holohan, P.J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, 20, 439-464.

Ministry of Economics and Finance (2007). *National programme of reforms*. Athens: Greek Ministry of Economics and Finance [in Greek].

Ministry of National Education and Religious Affairs (2005). *Higher education: Universities and technological educational institutions in Greece*. Athens: Ministry of National Education and Religious Affairs, Directorate of International Relations in Education.

Möller, I. (2002). The Ramsden Course Experience Questionnaire: A pilot study of final-year students on hospitality, leisure, sport and tourism degree courses.

Retrieved February 10, 2010, from:

<http://www.heacademy.ac.uk/assets/hlst/documents/projects/secq/ramsden.pdf>

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Oppenheim, A.N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Paulsen, M.B. (2002). Evaluating teaching performance. In C.L. Colbeck (Ed.), *Evaluating faculty performance* (pp. 5-18). San Francisco: Jossey-bass.

Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G., & Zepke, N. (2004). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: A synthesis of the research*. Wellington: Ministry of Education.

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

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

[doi:10.1080/03075079412331382143](https://doi.org/10.1080/03075079412331382143)

Richardson, J.T.E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment and Evaluation in Higher Education*, 30(4), 387-415. [doi:10.1080/02602930500099193](https://doi.org/10.1080/02602930500099193)

Schmitt, N.W., & Stults, D.M. (1985). Factors defined by negatively keyed items: The results of careless respondents? *Applied Psychological Measurements*, 9, 367-373. [doi:10.1177/014662168500900405](https://doi.org/10.1177/014662168500900405)

Spector, P. (2001). Research methods in industrial and organisational psychology: Data collection and data analysis with special consideration to international issues. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 10-26). London: Sage.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10. [doi:10.1007/BF02291170](https://doi.org/10.1007/BF02291170)

Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26. [doi:10.1177/014662168500900101](https://doi.org/10.1177/014662168500900101)

Woods, C.M. (2006). Careless responding to reverse-worded items:

Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186-191. doi:[10.1007/s10862-005-9004-7](https://doi.org/10.1007/s10862-005-9004-7)

Appendix A: Questionnaire items for the CEQ questionnaire

Teaching

- 3. *The teaching staff normally give me helpful feedback on how I am doing*
- 8. *The teaching staff of this course motivate me to do my best work*
- 16. *The staff make a real effort to understand difficulties I may be having with my work*
- 19. *My lecturers are extremely good at explaining things*
- 21. *The teaching staff work hard to make their subjects interesting*
- 27. *The staff put a lot of time into commenting on my work*

Skills Development

- 5. *The course has helped me develop my ability to work as a team member*
- 9. *The course has sharpened my analytical skills*
- 10. *As a result of my degree course, I feel confident about tackling unfamiliar problems*
- 18. *The course has developed my problem-solving skills*
- 22. *The course has improved my skills in written communication*
- 23. *My course has helped me to develop the ability to plan my own work*

Appropriate Assessment

- 13. *The staff seem more interested in testing what I have memorised than what I have understood*

17. *Feedback on my work is usually provided only in the form of marks or grades*

20. Too many staff ask me questions just about facts

26. *To do well in this course all you really need is a good memory*

Appropriate Workload

2. *There is a lot of pressure on me as a student in this course*

4. *The workload is too heavy*

15. *I am generally given enough time to understand the things I have to learn*

24. *The sheer volume of work to be got through in this course means it can't all be thoroughly comprehended*

Clear Goals and Standards

6. *I have usually had a clear idea of where I am going and what is expected of me in this course*

12. It is always easy to know the standard of work expected

25. *The staff made it clear right from the start what they expected from students*

29. *It has often been hard to discover what is expected of me in this course*

Academic Environment

1. The course is intellectually stimulating

7. The course administration is effective in supporting my learning

11. My course has stimulated my enthusiasm for further learning

14. Where it was used, information technology helped me to learn

28. I feel part of a group of students and staff committed to learning

30. I feel I benefit from being in contact with active researchers

Note: Number denotes item number on questionnaire, items in italics were retained in the final revised scales.

Table 1: Fit statistics for measurement models of CEQ

	CFI	TLI	RMSEA
Single common factor model	0.772	0.744	0.104
Two-factor model	0.886	0.893	0.060
Six-factor model	0.811	0.787	0.106
Revised five-factor model	0.954	0.943	0.042

Table 2: Coefficient Alpha values for revised CEQ scales

Scale	Cronbach's alpha
Teaching	0.85
Skills Development	0.88
Appropriate Assessment	0.82
Appropriate Workload	0.82
Clear Goals and Standards	0.84
Total scale reliability	0.88

Table 3: Relative importance of revised CEQ scales, as a percentage of R^2 , for predicting overall satisfaction

Scale	Percentage of R^2
Teaching	28.3%
Skills Development	37.2%
Appropriate Assessment	11.7%
Appropriate Workload	7.4%
Clear Goals and Standards	9.7%
Total % of variance explained	46.5%