

Putting student evaluations into perspective: The Course Experience Quality and Satisfaction Model (CEQS)

Abstract

Although not specifically designed for this purpose, the Course Experience Questionnaire (CEQ) continues to be used as a proxy for student satisfaction. This may be due to a lack of appropriate alternative measures, or a clear understanding of the relationship between quality and satisfaction. This study, therefore, examines the CEQ dimensions (*good teaching, clear goal setting, appropriate workload, appropriate assessment and generic skills development*) in conjunction with a *global student satisfaction* measure (developed for this study). In addition, the constructs of interest are conceptualised into a pattern of relationships that delineates antecedent and outcome variables in order to examine relationships. Data were collected via survey instrument from a sample of 164 Australian students attending one third-year capstone course of a Bachelor degree in business. The results indicate that *good teaching* and *clear goal setting* have a direct influence on outcome variables (i.e. *generic skills efficacy* and *overall satisfaction*), whereas *appropriate assessment* and *appropriate workload* do not have a direct effect on either of the outcome variables. However, the strongest influence in the hypothesised model was that exerted by *generic skills efficacy* on *overall satisfaction*. The structural model was further examined across male and female responses. While the male model replicated the findings of the overall model, the female model was quite different i.e. the antecedent variables directly influenced generic skills development but there was no direct influence on overall satisfaction.

Keywords

Program evaluation, evaluation utilization, CEQ, student satisfaction

Introduction

Over recent decades, higher education (HE) institutions have been under close scrutiny from both governments and consumers alike (Marsh, Morin, Ginns & Nagengast, 2011; Sumsion 2011; Waugh, 2003). Largely, this has been driven by initiatives born through government bodies (such as the Learning and Teaching Performance Fund in Australia) that reward those providers found to demonstrate excellence in teaching and learning (Davies, Hirschberg, Lye & Johnston, 2010). This has accentuated the need to monitor the quality of teaching and learning delivered by HE providers and, more so, ensures greater accountability, on the part of educators and institutions (Lyon & Hendry, 2002).

Of increasing importance has been the development and utilisation of performance indicators as a means to effectively measure and evaluate the quality of teaching and research outcomes (Byrne & Flood, 2003). Although limited measures have been made available to HE institutions to address the outcomes for evaluating the quality of teaching, one such measure, the Course Experience Questionnaire (CEQ), originating from a series of research activities conducted in the 1970s and 1980s, and further developed by Ramsden and Entwistle in Britain in the early 80s (Wilson, Lizzio & Ramsden, 1997), has been extensively used within HE environments. The Course Experience Questionnaire (CEQ) was created with the intent of measuring teaching effectiveness and perceived academic quality in relation to degree programs in HE institutions (Eley, 2001; Ramsden, 1991; Wilson et al., 1997). Numerous studies have investigated the reliability and validity of this intention (e.g., Byrne & Flood, 2003; Law & Meyer, 2011; Lizzio, Wilson & Simmons, 2002; Ramsden, 1991; Richardson, 1994; Trigwell & Prosser, 1991; Wilson et al., 1997).

In recent times, HE institutions and government bodies have been placing increasing importance on student satisfaction as a measure of the overall student HE experience (Gibson, 2010). Although the CEQ contains a single “satisfaction” item this is considered

insufficient to accurately measure this construct (Elliot & Shin, 2002). Moreover, the CEQ was not developed to be a proxy indicator of satisfaction (Ramsden, 2003) and the single satisfaction item was originally intended for use in scale validation (Ramsden, 1991). Despite this, there remains a misconception, largely driven by HE institutions and government bodies, regarding the relationship between the CEQ and student satisfaction (e.g., DEEWR, 2012; Griffith, 2012). The use of the CEQ, as a proxy for student satisfaction, may, in fact, occur due to the lack of appropriate alternative measures or a clear understanding of the relationship between quality and satisfaction.

This study, therefore, seeks to examine the relationships between perceived quality (as measured via the CEQ) and overall student satisfaction. In doing so, we identify antecedent and outcome variables, and the manner in which they impact upon each other, in order to further understand how overall student satisfaction comes to fruition. In addition, in acknowledging previous educational research in relation to gender differences (Grebennikov & Skaines, 2009; Richardson & King, 1991; Santhanam & Hicks, 2010; Centra & Gaubtz, 2000; Basow, Phelan & Capotosto, 2006; Snow, 1977), we further examine and compare these relationships across male and female students. The results provide significant implications for both theory and practice and a solid platform for future research in this important area of research.

Course Experience Questionnaire (CEQ)

Originally based on a series of research activities conducted in the 1970s and 1980s (Wilson et al., 1997), the CEQ was developed in Australia as a means of collecting information about graduates' quality perceptions of their degree program upon completion (Eley, 2001; Ramsden, 1991). Since its development, the CEQ has been utilised as an annual survey of graduates, conducted through HE institutions (Coates, 2007; Wilson et al., 1997) and by the

Graduate Careers Council of Australia (Eley, 2001). The CEQ was implemented to achieve two primary objectives. Firstly, the CEQ was used as a means of maintaining ‘assurance and accountability’ (Griffin et al. 2003, p. 260) through providing HE institutions with a mechanism to analyse student ratings of their degree programs so as to develop performance indicators of quality teaching (Griffin et al., 2003, Ramsden, 1991). Secondly, comparative results of the CEQ assisted HE institutions in the development of more sound processes to enhance the level of teaching quality (Griffin et al., 2003). On initial development of the CEQ, the key purpose was to address five set criteria: (1) obtain accurate opinions from students in relation to the main areas associated with quality teaching and curriculum, (2) test and validate the scale items, (3) develop an economical viable and reliable survey instrument, (4) ensure the cross-utilisation of the survey to all Australian HE institutions and, (5) ensure that the survey could determine differences between student perceptions in relation to different academic courses (Ramsden, 1991).

Although Ramsden’s (1991) research began with an 80-item scale, a smaller scale was always the intention and, as such, there have been several CEQ versions, for example, containing 36 items, 30 items, 25 items and 23 items (e.g., Eley, 2001; Lyon & Hendry, 2002; Wilson et al., 1997). Whilst, there are, to date, several versions of the instrument, the 25-item survey remains the most commonly used variant of the CEQ (Lyon & Hendry, 2002; McInnis, Griffin, Richard & Coates, 2001; Waugh, 1998). The 25-item CEQ version comprises five scale dimensions including, *good teaching* (6 items), *clear goals and standards* (5 items), *appropriate workload* (4 items), *appropriate assessment* (3 items), and *generic skills* (6 items) (McInnis et al., 2001) and a single item addressing *overall satisfaction* with the quality of the course (McInnis et al., 2001).

Distinguishing course quality and student satisfaction

Student satisfaction is an important research area as several studies have drawn a link between student satisfaction and student motivation, student retention and recruiting efforts in HE settings (Douglas, Douglas & Barnes, 2006; Elliot & Healy, 2001; Elliot & Shin, 2002; Hoyt, 1973). Moreover, student satisfaction is commonly pronounced within university mission statements, goals and objectives, marketing strategies and promotional activities (Elliot & Shin, 2002). To this end, institutions are acknowledging that in order to retain students, they are required to re-engineer the focus of HE institutions to continuously develop strategies that, at minimum, meet and, at best, exceed student expectations (Cheng & Tam 1997; De Shields, Kara & Kaynak, 2005; Elliot & Healy, 2001; Elliot & Shin, 2002).

Applying marketing theory in the context of HE is appropriate as, essentially, students are the HE institution's customers (Douglas et al., 2006), in that students exchange money in return for an education (product). As the fight for the student dollar heightens, HE institutions need to be mindful of the role their customers (i.e. students) play as co-marketers of the institution. For example, the power of word-of-mouth communications is well documented (e.g., Dibb & Simkin, 1993; Krishnan & Hartline, 2001 etc.) and increasing, due to the limitless geographic boundaries resulting from electronic communications. While it is important for HE institutions to continue to monitor the quality of their programmes (from a pedagogical perspective), it is also important that they acknowledge the perception of quality and how this impacts on satisfaction from the student's (customer's) perspective. To this end, it is important to clearly delineate, in the HE setting, the concepts of course quality and satisfaction.

Satisfaction has been conceptualized in many different ways. For example, Tanner, Jr (1996) argues that overall satisfaction is comprised of both product satisfaction and process satisfaction. Woodruff and Schumann (1993), on the other hand, advocate that satisfaction is

an immediate reaction to how much value is received from using the product in specific use situations. In addition, Spreng, MacKenzie and Olshavsky (1996) promote overall satisfaction as product attribute satisfaction and information satisfaction, while Crosby and Stephens (1987) argue that overall satisfaction is comprised of satisfaction with contact personnel, core service and the institution. Notwithstanding their differences, in all the proposed satisfaction frameworks there are two common elements.

Firstly, there is general agreement that satisfaction is an emotive variable, which encapsulates an individual's evaluative response to both product performance and non-product performance (Tanner, 1996; Spreng et al., 1996; Oliver, 1993). In the educational setting, perceived quality related to product performance (i.e. course/program performance) includes student perceptions of teaching, standard setting, assessment techniques and workload etc. From a non-quality perspective, students' perception of other service aspects, such as timetabling, facilities and technology, administrative processes, ease of parking or access to public transport, and in more recent years, flexibility of delivery mode will also impact on the students' overall satisfaction. Therefore, quality perceptions of product and non-product performance *precede* satisfaction. In fact, this notion is supported by the second common school of thought i.e. that satisfaction is an outcome variable of service quality perceptions (Cronin, Brady & Hult, 2000; Hallowell, 1996; Fornell et al., 1996; Athansopoulos, 2000). In fact, Bagozzi (1992) argues that the perception of service quality is an appraisal (or cognitive response) which leads to satisfaction, which is an emotional response, that in turn drives behavior.

Thus, a clear distinction between course quality (antecedent variable) and student satisfaction (outcome variable) is made. Course quality relates to the students perception of all aspects related specifically to the course/program, whereas overall satisfaction is the student's evaluative response to both course-related (i.e. product) and non-course-related (i.e.

non-product). Moreover, for the purposes of this study, student satisfaction is defined as the student's emotional and behavioral response to course appraisal and is operationalized accordingly.

Common criticisms of the CEQ are embedded in the important distinction between teaching/course quality and student satisfaction. For example, it is argued that the CEQ does not embrace sufficient scales (e.g., Griffin et al., 2003; McInnis et al., 2001; Sheridan, 1995) to provide a complete picture of a student's HE experience (e.g., availability of computers, library facilities) (Douglas et al., 2006; Price et al., 2003) which limits clarity of the actual factors contributing to student perceptions of quality. In other words, the CEQ is criticised for not measuring what is previously referred to as "non-product" aspects of the service experience. In addition, as the CEQ is normally administered at the completion of a degree programme, the resultant lack of timely results inhibits HE institutions from implementing programmes to assist faculty staff in improving the quality of teaching outcomes (Davies et al., 2010). In actual fact, the CEQ was designed to measure perceived *quality* (not satisfaction) and the individual satisfaction item (included in the 25-item CEQ version) was merely an "add-on" to the original scale (Richardson, 2009).

We argue here that, while the CEQ is a validated measure of course/program quality that is multi-dimensional, it is distinctly different from overall student satisfaction. On this basis, we explore the dimensions measured by the CEQ (i.e. good teaching, clear goal and standard setting, appropriate assessment and appropriate workload) within the realm of a network of relationships that explore the influence of course evaluation (as measured by CEQ) on student satisfaction. On this basis we propose the Course Experience and Satisfaction Model (CEQS) as a conceptual mechanism by which to undertake this investigation and, as such, this model is further discussed.

Conceptual development: Course Experience Quality and Satisfaction Model (CEQS)

In order to examine the relationships between satisfaction and quality in HE, we adopt the well-documented and validated framework of quality dimensions represented within the CEQ. As depicted in Figure 1, we have grouped the CEQ six dimensions according to, firstly, their sequential positioning within the network of relationships i.e. antecedents and outcomes. Clearly, when examining the questionnaire items, two dimensions are identified as outcome variables and these include *overall satisfaction* and *generic skills efficacy*. These are variables that occur as a *result* of the antecedent dimensions, shown as *good teaching, clear goal setting, appropriate assessment and workload*. In other words, satisfaction and generic skills efficacy are dependent upon the delivery of quality teaching, clear goal setting and perceptions of appropriate assessment methods and workload.

Hypotheses Development

In order to develop the hypotheses of this study, we examine the individual dimensions within their respective categories (as depicted in Figure 1). In terms of *overall satisfaction* and *generic skills efficacy*, being the key dependent variables of the CEQS Model, we expect that these two will be related. In this relationship, we model generic skills efficacy as the antecedent to overall satisfaction. For example, if one believes they have gained good generic skills through undertaking a particular course (or degree), then they are likely, as a result, to express overall satisfaction. On this basis we propose:

H1: Student generic skills efficacy scores will have a significant positive influence on overall course satisfaction scores.

Given that the rationale behind the CEQS is to measure perceived quality and satisfaction, it is expected that the student's experience with the course in the form of teaching quality and the clear communication of goals and standards will influence both their generic skills efficacy and overall satisfaction. For example, if the student has had a good learning experience, where learning objectives are clearly understood and the achievement of same has occurred through good teaching, then it is likely that they will believe, as a consequence, their skill base has been enhanced and overall satisfaction with the course will result. Thus, we hypothesise that:

H2: Student evaluation of good teaching scores will have a significant positive direct influence on both:

(a) generic skills efficacy scores

(b) overall course satisfaction scores

H3: Student evaluation of clear goals/standards scores will have a significant positive direct influence on both:

(a) generic skills efficacy scores

(b) overall course satisfaction scores

H4: Perceived appropriateness of assessment scores will have a significant positive direct influence on both:

(a) generic skills efficacy scores

(b) overall course satisfaction scores

H5: Perceived appropriateness of workload scores will have a significant positive direct influence on both:

(a) generic skills efficacy scores

(b) overall course satisfaction scores

Gender Differences:

Finally, our testing of the proposed CEQS Model will not be complete if we do not consider potential group differences in examining our model. Clearly the literature provides strong documentation of research that has consistently provides evidence of significant differences between female and male students (Al-Harthi, 2005; Grebennikov & Skaines, 2009; Richardson & King, 1991). For example, female students are found to be more receptive to learning behaviours and educational discourses (Smith, 2004), are more likely to conform and be conscientious (Chapman, 1996; Rudd, 1984), have higher attendance rates (Woodfield, Jessop & McMillan, 2006) and, as a result, enjoy a more collaborative connected learning environment (Blum, 1999). On the downside, female students have less self-confidence (Sadler-Smith, 1996) are more anxious and have a higher fear of failure (Osbourne, 2001). These findings are important in terms of the CEQS Model as application to study and course participation has been found to significantly affect student perceptions of both teaching and course quality (Burns & Ludlow, 2006; Voelkl, 1995). Although there have been mixed findings in relation to gender and student evaluations, Basow, Phelan and Capotosto (2006) found that male students tend to rate female faculty lower than female students and this is supported by earlier findings of Centra and Gaubatz (2000). In addition, Santhanam and Hicks (2002) found that female students tended to rate their professors higher than male students, regardless of the gender of the professors. They attributed this to different learning style across gender. However, in contrast, Grebennikov and Skaines (2009) found

that, while female students have higher expectations, they have lower satisfaction levels than male students (Grebennikov & Skaines, 2009). Given these mixed findings, we suggest that the manner in which female and male students respond to the CEQ will exhibit statistically different patterns. This being the case, we hypothesise the following:

H6: There will be a difference in the structural model (CEQ Network) across male and female students.

Research design

The commonly-used 25-item Course Experience Questionnaire (CEQ) (Lyon & Hendry, 2002; McInnis et al., 2001; Waugh, 1998) provided the framework for survey development. Therefore, *good teaching* (GTS) was measured by 6 items, *clear goals and standards* (CGS) by 5 items, *appropriate assessment* (AAS) by 3 items, *appropriate workload* (AWS) by 4 items and *generic skills efficacy* (GSS) by 6 items. Given that the CEQ only includes one item to measure *overall satisfaction* (OS) and that this construct is a key dependent variable in this study, we developed a further four satisfaction items to capture other aspects of satisfaction such as intention to recommend, emotion (enjoyableness), ranking compared to other courses and need (or no need) for improvement. This resulted in five items measuring overall satisfaction (shown below):

- *Overall, I am satisfied with the quality of this course (CEQ item).*
- *Overall, I would recommend this course to others (developed for this study).*
- *Overall, this is one of the best courses I have taken (developed for this study).*
- *Overall, this course does not need improvement (developed for this study).*
- *Overall, I enjoyed taking this course (developed for this study).*

In addition, demographic variables such as *gender, age, enrolment type, nationality* and *hours of paid work* were included for the purpose of group classification and sample description.

Data collection

Data was collected via a survey administered to third-year marketing students undertaking the capstone course for the Bachelor of Business degree at an Australian university. The decision to restrict data collection to one course was to control for intervening variables that may exist if data were collected across various courses or degrees. In other words, we wanted students to have the same frame of reference and be in the same stage of their degree. While this decision did restrict the generalizability of the findings, we argue that this study is exploratory in nature and requires further validation in a wider context. Students were instructed to answer the survey questions in relation to their current third-year course and data was collected at the completion of the final lecture for the course. The survey was distributed to a convenience sample of 210 students, which resulted in 164 completed surveys being returned, thus representing a response rate of 78%.

The sample

The sample comprised of 47% male and 52% female. Ages ranged from 21 years to 32 years and the mean age was 25 years (standard deviation of 3.7). Ninety percent of students were full-time enrolled while the remaining 10% were part-time students. Number of paid working hours per week ranged from 0 to 40, with a mean of 16 hours per week (standard deviation of 5.2). Predominantly, the sample comprised of Australian students (54%), followed by Canadian students (18%), Swedish students (11%), New Zealand students (4%). The remaining students came from various countries within Europe, USA and Asia.

Data Preparation

Firstly, the completeness of the data set was examined. Very few missing data points (i.e., 17) were identified and no systematic patterns of missing data was evident. On this basis, it was appropriate for missing data to be replaced by linear interpolation estimation (Howell, 2009) using SPSS. Secondly, the normality of variables was inspected via measures of central tendency and dispersion and visual inspection of histograms. There were no obvious anomalies detected at this point.

Analysis and Results

Careful consideration was given to the most appropriate analysis in order to effectively test the hypotheses of this study, given that, comparative groups of males and females were relatively small. The chosen method was Partial Least Squares (PLS) analysis undertaken in the statistical package SmartPLS (Ringle, Wende & Will, 2005). PLS is a multivariate analysis technique allowing the empirical assessment of a structural model in conjunction with its measurement model (Wold, 1982). At the conceptual core of PLS, and guided by theory, is a combination of principal components analysis (Vandenberg, 1996), relating measures to constructs, and path analysis (Bontis, 1998), relating to causal links between constructs. The objective in PLS is to maximise the explanation variance and, therefore, the model is considered to perform well if the R^2 and the relationships among constructs are significant (Bontis, 1998). Importantly, PLS calculates parameters using least squares estimations and is able to generate useful results with small samples (Tiessen & Linton, 2000).

Measurement model results (outer model):

Examination of the measure model revealed that some items did not load significantly on their respective latent constructs. Thus, items exhibiting t-values < 1.96 (two-tailed test)

were deleted from further analysis (Chin, 1998). On this basis, it was necessary to delete the following six items: 1 item measuring *good teaching* (GTS1), 1 item measuring *clear goals and standards* (CGS5), 1 item measuring *appropriate assessment* (AAS1), 2 items measuring *appropriate workload* (AWS1 and AWS4) and 1 item measuring *generic skills* (GSS3). Re-running the analysis, all remaining items loaded significantly (t-values in excess of 1.96) with factor loadings ranging from .63 to .94. Construct reliabilities were all above the recommended level of .70 (Bagozzi & Yi, 1988) and ranged from .83 to .94. Furthermore the average variance extracted for each construct was greater than the acceptable level of .50 (Fornell & Larcker, 1981). (Refer to Table 1 for details for the measurement model analysis).

---- **Insert Table 1 here** ----

Validity checks

Prior to hypotheses testing, it is important to determine the integrity of the data through establishing evidence of construct validity. Construct validity is established through the examination of convergent and discriminant validity. *Convergent validity* is determined when variables that are theoretically expected to be similar, actually are correlated, while *discriminant validity* is established when variables that are theoretically expected to be different, are not correlated (De Vellis, 2011). Convergent validity is evaluated by the strength and significance of the loadings, the average variance extracted (AVE) and the construct reliability (Bagozzi & Heatherton, 1994). Given that all factor loadings were significant and strong ($t > 1.96$) (Gerbing & Anderson, 1988), the AVE greater than .50 (Fornell & Larcker, 1981) and construct reliabilities greater than .70 (Bagozzi & Yi, 1988), then convergent validity is established (refer Table 3). Discriminant validity is evaluated by Fornell and Larcker's (1981) criterion whereby discriminant validity is said to be evidenced if

the average variance extracted (AVE) is greater than the square of the construct's correlations with the other factors. As all AVEs were greater than the squared construct correlations, evidence of discriminant validity was established (refer Table 2).

---- **Insert Table 2 here** ----

Structural model results (hypothesis testing)

Analysis of the structural model was undertaken via PLS and the results, as shown in Figure 1, indicate that five paths in the model were significant, having critical ratios which exceeded the two-tailed testing cut-off of 1.96. They were GSS/OS; GTS/GGS; GTS/OS; CGS/GGS and CGS/OS thus providing evidence of the hypothesised relationships in H1, H2 and H3. Variables labelled as AAS and AWS had no significant influence on either of the key dependent variables, GSS and OS. Thus, there was no evidence of relationships as hypothesised in H4 and H5. R^2 values relating to GGS (.48) and SATIS (.40) were strong and significant (i.e., above the recommended level of .10 as recommended by Falk and Miller (1992). The overall average variance accounted for in the model was also strong at .44.

---- **Insert Figure 1 here** ----

Gender comparisons

In order to address H6 and make appropriate gender comparisons, the sample was split across male and female respondents. This resulted in a relatively even split with Model A (Males) comprising of 78 and the number of females respondents (Model B) being 86. Analyses of these models were run separately in SmartPLS and the results appear in Figure 2. The male model (Model A) performed identically, in terms of the significance of paths, as did the

overall model. That is, significant paths included GSS/OS; GTS/GGS; GTS/OS; CGS/GGS and CGS/OS. R-squared values were strong for GSS (.53) and OS (.56). Thus, the average variance accounted for across this model was strong at .55. The female model (Model B), however, did reveal different relationships than did Model A, thus providing evidence for H6. For example, in the female model, there were no significant direct influence between the antecedent variables (GTS, CGS, AAS and AWS) and overall satisfaction (OS). The only significant influence on OS in Model B was GSS. In terms of the relationships between the antecedent variables and GSS, there were three significant influences found i.e. from GTS, CGS and AWS. R-squared values were also significant with GSS (.38) and OS (.46) and the average variance accounted for in the model was .42. For ease of comparisons, both models including statistical results appear in Figure 2.

---- **Insert Figure 2 here** ----

Discussion

The findings of this study highlight the importance of examining the relationships between the CEQ dimensions, in order to establish the “sequential” process by which student satisfaction is achieved. The CEQS Model is proposed as a mechanism for use at the completion of courses or degrees, just as the CEQ was proposed (Eley, 2001; Ramsden, 1991; Wilson et al., 1997). The results of model testing overwhelmingly demonstrate that the key antecedents to the subsequent assessment of student generic skill development and overall satisfaction are good teaching practises combined with the explicit expression of course learning goals and standards. Whether the student believes the associated assessment and workload is appropriate, or not, has no direct impact on their perceived generic skill development or satisfaction.

Appropriate assessment (AAS) and workload (AWS) measures

An interesting finding in the overall model was that appropriate assessment (AAS) and appropriate workload (AWS) had no influence on generic skills efficacy (GSE) and overall satisfaction (OS). We propose the following as a possible explanation for this finding. It may be that students find it extremely difficult to determine what is *appropriate* assessment and workload as they lack a valid frame of reference. While some students may respond on the basis of appropriateness, in relation to their learning, others may respond on the basis of appropriateness in relation to their social and work commitments or, further, on the basis of appropriateness in relation to things they like or dislike doing. On this basis, it is not surprising that, due to the variability in responses in relation AAS and AWS, relationships with other more definitive measures (e.g., GTS, CGS, GSE and OS) have not been found. As such, the findings do not suggest that the CEQ is effectively obtaining accurate opinions from students in relation to the main areas associated with quality teaching and curriculum, as was its original aim (Ramsden, 1991). Therefore, it is very troubling, indeed, as pedagogically sound course assessment and workload may well be altered on the basis of survey responses to questions that do not accurately reflect reality. This could well be a major concern for those institutions using the CEQ scale as a means by which to modify teaching in practice.

While it could be easily argued that the statistical validity (construct, convergent, discriminant) of the AAS and AWS scale items, and even the face validity of these items, has been demonstrated on numerous occasions (e.g., Broomfield & Bligh, 1998; Byrne & Flood, 2003; Byrne & Willis, 2004; Lyon & Hendry, 2002; Trigwell & Prosser, 1991; Waugh 1998, etc.), this is irrelevant because in developing these items there has been little regard for who will actually respond to them (i.e. students). Thus, with the perspective of students (i.e.

student knowledge and frames of reference) clearly in mind, the measures associated with AAS and AWS require re-development if meaningful results are to be obtained. We concede that this is conjecture at this point, however, the findings do indicate that more research is needed to investigate possible response bias in relation to these dimensions.

Generic skills efficacy (GSS)

Generic skills efficacy or, in other words, the belief students have that their generic skills have been enhanced, plays a pivot role in the tested conceptual model. Not only does good teaching (GTS) and clear goal setting (CGS) directly impact on GSS, but GSS also has a direct impact on student satisfaction. The key role of GSS in the model is further exemplified when gender comparisons are made. For female students, GSS is the *only* construct that directly affects satisfaction. In fact, GSS fully intervenes the pathway between the antecedent variables (i.e. GTS and CGS) and satisfaction. For male students, however, there are direct paths to both GSS and satisfaction. This result confirms the findings of Grebennikov and Skaines (2009) who also found differences based on student expectations and satisfaction across gender.

To explain this finding, we draw on the research of Sadler-Smith (1996) who found that female students were less self-confident in their academic ability, reported higher levels of anxiety and had a greater fear of failure in relation to their studies. On this basis, we argue that, as female students lack self confidence in their learning, for satisfaction to occur, they must have confidence (or a belief) that their generic skills have been enhanced. The satisfaction of male students, however, is not totally dependent on this belief. Given male students increased level of self-confidence in this context, it could even be possible that they believe they have the skills when they do not, and this could be due to over-confidence in

their own ability. Gender differences aside, GSS is important for both male and female students and this needs to be given careful consideration in course development and delivery.

Recommendation for course development and delivery

The findings have significant implications for course planning and classroom delivery. The fact that good teaching impacts on student satisfaction and generic skill development is not really surprising, and doesn't present us with anything more than what we already suspect. However, the key implications of this study reside under the umbrella label of *explicitness* i.e. firstly; being explicit in what learning is trying to be achieved by the course and secondly; being explicit in what students have gained by the course (i.e. generic skills development). It is not good enough to simply state course objectives and outcomes in the course outline as these statements could be misunderstood by students, or not even read for that matter. It is not good enough to state learning objectives at the commencement of each lecture as students may have difficulty in placing these objectives in relation to the course objectives and subsequent skill development. What is required is considerable classroom time being devoted to continually *telling* students what needs to be achieved, what is required for them to achieve, and how their course achievements will have enhanced their skill base. Academics might be surprised, when thinking through this approach, to find that past teaching and assessment practices may reveal themselves as being irrelevant or hard to substantiate as effective learning mechanisms.

By adopting teaching and assessment practices that can be *explicitly* linked to both objectives and outcomes (and discarding those that cannot), students will be motivated to "buy in" to the course because they understand what they are doing, why they are doing it and what they will get from it in the end. The latter awareness (i.e. what they get) is particularly important for female students. This is so because, in order for females to be

confident that their skills enhancement, this must be explicitly reinforced to them, not only at course end, but at appropriate developmental stages throughout. The treatment of all students in this manner, regardless of gender, will no doubt produce positive outcomes. Using this approach will result in incremental skill development through the degree program and result in students having a well-rounded appreciation of their skill development at degree end. This is important as students should not be *surprised* by what they can do when they enter the “real world”, they must *know* what they can do before they leave university.

Limitations:

The survey methodology used to collect the data for this study is acknowledged as a limitation. This method is reliant on respondents being able to accurately report their level of agreement with statements that delve into their attitudes, feelings and perceptions, and, as result, measurement error is unavoidable. However, data analysis of the measurement model suggests that measurement error was not problematic and this is reflected in the reliability analysis and other validity checks. Another limitation of data collection occurs when sample size is small. While the analytical procedure adopted (i.e. PLS) specifically accommodates small sample sizes, the representativeness of the sample may be questionable. Furthermore, as data were collected from in one higher education institution in Australia the generalizability of the results may well be limited. This being the case, we concede that this research is largely exploratory and requires further validation in a wider context.

Future Research

The CEQS Model is posed as a valid means by which to understand the relationships between course quality aspects and the associated outcomes. As previously argued, there needs to be further scale development in relation to the constructs that group under the CEQS label of

“perceived appropriateness”. Not only has this study found these items to be problematic, but this has also been evident in other studies (e.g., Curtis & Keeves, 2000; Waugh, 1998). This would enhance the validity of the CEQS, as a whole, and set the scene for further testing. As the CEQS Model is proposed as a course, or degree level, evaluation tool, future research could benefit from its use across students at different stages throughout their degree (between group design) or track individual student progress in a longitudinal fashion (within group design). Furthermore, as gender comparisons in this study proved fruitful in providing important information about what influences the course evaluations of males and females so, too, can future group comparisons (e.g., between international versus domestic students, on-line versus on-campus students, full-time versus part-time students, etc.). In summary, more work on the CEQ scale items is warranted. However, the result may well eventuate in a very useful measurement tool that can effectively gather performance and evaluation data as students progress through their programs with the view to achieving positive post-graduate outcomes.

Conclusion

While considerable research has focussed on the CEQ, this study adapts its traditional use to that which examines student course experiences (i.e. perceived quality and overall satisfaction) through a nomological lens (system of relationships). In doing so, significant insight has been provided in terms of the important influences associated with student perceived outcomes. The CEQS Model, proposed and tested within this study, highlights the need for further scale development in order to develop a sound course evaluation mechanism that can be used at the course and program level. The findings highlight that it is not *what* students evaluate, but *how* students evaluate courses that is the key to improved pedagogical performance.

References

- Al-Harhi, A. S. (2005). Distance experiences of arab gulf students in the united states: A cultural perspective. *The International Review of Research in Open and Distance Learning*, 6(3), 1-14.
- Bagozzi, R. P. & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(1), 35-67.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), pp.74-94.
- Blum, K. D. (1999). Gender differences in asynchronous learning in higher education: Learning styles, participation barriers and communication patterns. *Journal of Asynchronous Learning Network*, 3(1), 1-21.
- Bontis, N. (1998). Intellectual capital: An exploratory study that develops measures and models. *Management Decision*, 36(2), 63-76.
- Broomfield, D., & Bligh, J. (1998). An evaluation of the 'short-form' course of experience questionnaire with medical students. *Medical Education*, 32(4), 367-369.
- Burns, S. M., & Ludlow, L. H. (2006). Understanding student evaluations of teaching quality: The contributions of class attendance. *Journal of Personnel Evaluation in Education*, 18(2), 127-138.
- Byrne, M., & Flood, B. (2003). Assessing the teaching quality of accounting programmes: An evaluation of the course experience questionnaire. *Assessment & Evaluation in Higher Education*, 28(2), 135-145.
- Byrne, M., & Willis, P. (2004). Leaving certificate accounting: Measuring students' perceptions with the course experience questionnaire. *Irish Educational Studies*, 23(1), 49-64.
- Chapman, K. (1996). An analysis of degree results in geography by gender. *Assessment & Evaluation in Higher Education*, 21(4), 293-313.
- Cheng, Y. C., & Tam, W. M. (1997). Multi-models of quality in education. *Quality Assurance in Education*, 5(1), 23-31.
- Chin, Wynne (1998)**
- Coates, H. (2007). Excellent measures precede measures of excellence. *Journal of Higher Education Policy and Management*, 29(1), 87-94.
- Crosby, L.A., & Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry. *Journal of Marketing Research*, 24(4), 404-411.
- Curtis, D.D., & Keeves, J.P. (2000). The course experience questionnaire as an institutional performance indicator. *International Education Journal*, 1(2), 73-82.
- Davies, M., Hirschberg, J., Lye, J., & Johnston, C. (2010). A systematic analysis of quality of teaching surveys. *Assessment & Evaluation in Higher Education*, 35(1), 87-100.
- Department of Education, Employment and Workplace Relations. (2011). *Performance funding: Administrative guidelines*. Retrieved from <http://www.deewr.gov.au/highereducation/policy/documents/administrativeguidelines.docx>
- De Shields, O. W., Jr., Kara, A., & Kaynak, E. (2005). Determinants of business student satisfaction and retention in higher education: Applying herzberg's two factor theory. *International Journal of Educational Management*, 19(2), 128-139.
- De Vellis, ?? (2011)**
- Dibb, S., & Simkin, L. (1993). The strength of branding and positioning in services. *International Journal of Service Industry Management*, 4(1), 25-35.

- Douglas, O. W., Douglas, A., & Barnes, B. (2006). Measuring student satisfaction at a UK university. *Quality Assurance in Education, 14*(3), 251-267.
- Eley, M. (2001). The course experience questionnaire: Altering question format and phrasing could improve the ceq's effectiveness. *Higher Education Research & Development, 20*(3), 293-312.
- Elliot, K. M., & Healy, M. A. (2001). Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education, 10*(4), 1-11.
- Elliot, K. M., & Shin, D. (2002). Student satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management, 24*(2), 197-209.
- Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. Akron, OH: University of Akron Press.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25*(2), 186-192.
- Gibson, A. (2010). Measuring business student satisfaction: A review and summary of the major predictors. *Journal of Higher Education Policy and Management, 32*(3), 251-259.
- Grebennikov, L., & Skaines, I. (2009). Gender and higher education experience: A case study. *Higher Education Research & Development, 28*(1), 71-84.
- Griffin, P., Coates, H., Mcinnis, C., & James, R. (2003). The development of an extended course experience questionnaire. *Quality in Higher Education, 9*(3), 259-266.
- Griffith University. (2012). *Australian Graduate Survey*. Retrieved from <http://www.griffith.edu.au/student-surveys/australian-graduate-survey>.
- Howell, D. (2009)
- Hoyt, D. P. (1973). Measurement of instructional effectiveness. *Research in Higher Education, 1*(4), 367-378.
- Krishnan, B.C., & Hartline, M.D. (2001). Brand equity: Is it more important in services? *Journal of Services Marketing, 15*(5), 328-342.
- Lizzio, A., Wilson, K., & Simmons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*(1), 27-51.
- 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 & Evaluation in Higher Education, 27*(4), 339-351.
- Marsh, H.W., Ginns, P., Morin, A.J., & Nagengast, B. (2011). Use of response ratings to benchmark universities: Multilevel modeling of responses to the Australian course experience questionnaire (CEQ). *Journal of Educational Psychology, 103*(3), 733-748.
- Mcinnis, C., Griffin, P., Richard, J., & Coates, H. (2001). *Development of the course experience questionnaire (ceq)*. Melbourne: Report for Department of Education, Training and Youth Affairs (Melbourne).
- Osbourne, J. W. (2001). Testing stereotyping threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26*(3), 291-310.
- Price, I. F., Matzdorf, F., Smith, L., & Agahi, H. (2003). The impact of facilities on student choice of university. *Facilities, 21*(10), 212-222.

- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education, 16*(2), 129-149.
- Ramsden, P. (2003). Student surveys and quality assurance. In C. S. Nair & R. Harris (Eds.) *National Quality in a Global Context* (pp. 121-130). Melbourne: Australian Universities Quality Agency.
- Richardson, J.T.E. (1994). A British evaluation of the course experience questionnaire. *Studies in Higher Education, 19*(1), 59-68.
- Richardson, J.T.E. (2009). What can students' perception of academic quality tell us? Research using the course experience questionnaire. In M. Tight, K.H. Moh, J. Huisman & C.C. Morphew (Eds.), *The Routledge International Handbook of Higher Education* (pp. 199-209). New York: Routledge.
- Richardson, J. T. E., & King, E. (1991). Gender differences in the experience of higher education: Quantitative and qualitative approaches. *Educational Psychology, 11*(3/4), 363-382.
- Ringle, C. M., Wende, S., & Will, A. (2005). *Smartpls 2.0 (beta)*. Retrieved from <http://www.smartpls.de>.
- Rudd, E. (1984). A comparison between the results achieved by women and men studying for first degrees in british universities. *Studies in Higher Education, 9*(1), 47-57.
- Sadler-Smith, E. (1996). Approaches to studying age, gender and academic performance. *Educational Studies, 22*(3), 367-379.
- Sheridan, B. (1995). *The course experience questionnaire as a measure for evaluating courses in higher education*, Perth: Edith Cowan University.
- Smith, F. (2004). 'it's not all about grades'. Accounting for gendered degree results in geography at brunel university. *Journal of Geography in Higher Education, 28*(2), 167-178.
- Spreng, R.A., MacKenzie, S.B., & Olshavsky, R.W. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of Marketing, 60*(3), 15-32.
- Sumsion, J. (2011). Capacity building in earlt childhood education research in a regional Australian university. *British Journal of Educational Studies 59*(3), 265-284.
- Tanner, J.F., Jr. (1996). Buyer perceptions of the purchase process and its effect on customer satisfaction. *Industrial Marketing Management, 25*(2), 125-133.
- Tiessen, J. H., & Linton, J. D. (2000). The jv dilemma: Cooperating and competing in joint ventures. *Canadian Journal of Administrative Sciences, 17*(3), 203-216.
- Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: The influence of learning context and student approaches to learning outcomes. *Higher Education, 22*(3), 251-266.
- Vandenbosch, M. B. (1996). Confirmatory compositional approaches to the development of product spaces. *European Journal of Marketing, 30*(3), 23-46.
- Voelkl, K. E. (1995). School warmth, student participation, and achievement. *The Journal of Experiemental Education, 63*(2), 127-138.
- Waugh, R. F. (1998). The course experience questionnaire: A rasch measurement model analysis. *Higher Education Research & Development, 17*(1), 45-64.
- Waugh, R.F. (2003). Evaluation of quality student experiences at a university using a rasch measurement model. *Studies in Educational Evaluation, 29*(2), 145-168.
- Wilson, K. L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the course experience questionnaire. *Studies in Higher Education, 22*(1), 33-53.
- Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Joreskog & H. O. A. Wold (Eds.) *Systems under indirect observations*. (pp. 1-54). Amsterdam: North-Holland.

Woodfield, R., Jessop, D., & Mcmillan, L. (2006). Gender differences in undergraduate attendance rates. *Studies in Higher Education, 31*(1), 1-22.

Woodruff, R.B., Schumann, D.W., & Gardial, S.F. (1993). Understanding value and satisfaction from the customer's point of view. *Survey of Business, 29*(1), 33-40.

Table 1: Results of the Measurement Model

Items/Constructs	Factor Loading	t-value	Composite Reliability	Average Variance Extracted
Good Teaching (GTS)				
GTS2	.71	9.57	.85	.53
GTS3	.75	11.42		
GTS4	.76	10.98		
GTS5	.74	12.19		
GTS6	.67	7.67		
Clear Goals/Standards (CGS)				
CGS1	.74	10.26	.83	.55
CGS2	.80	18.56		
CGS3	.63	6.01		
CGS4	.78	11.42		
Appropriate Assessment (AAS)				
AAS2	.67	3.14	.87	.67
AAS3	.94	5.65		
Appropriate Workload (AWS)				
AWS2	.94	6.53	.87	.67
AWS3	.66	3.01		
Generic Skills Efficacy (GSS)				
GSS1	.74	13.52	.84	.52
GSS2	.71	10.01		
GSS4	.75	11.28		
GSS5	.70	8.84		
GSS6	.70	8.96		
Overall Satisfaction (OS)				
OS1	.84	24.37	.94	.66
OS2	.80	39.53		
OS3	.84	19.28		
OS4	.86	36.16		
OS5	.70	8.54		

Table 2: Discriminant Validity Results

	GTS	CGS	AAS	AWS	GSS	OS
GTS	.53	.48	.36	.06	.54	.30
CGS		.55	.32	.09	.48	.32
AAS			.67	.12	.17	.15
AWS				.67	.03	.19
GSS					.52	.32
OS						.66

Figure 1: Overall Model Results

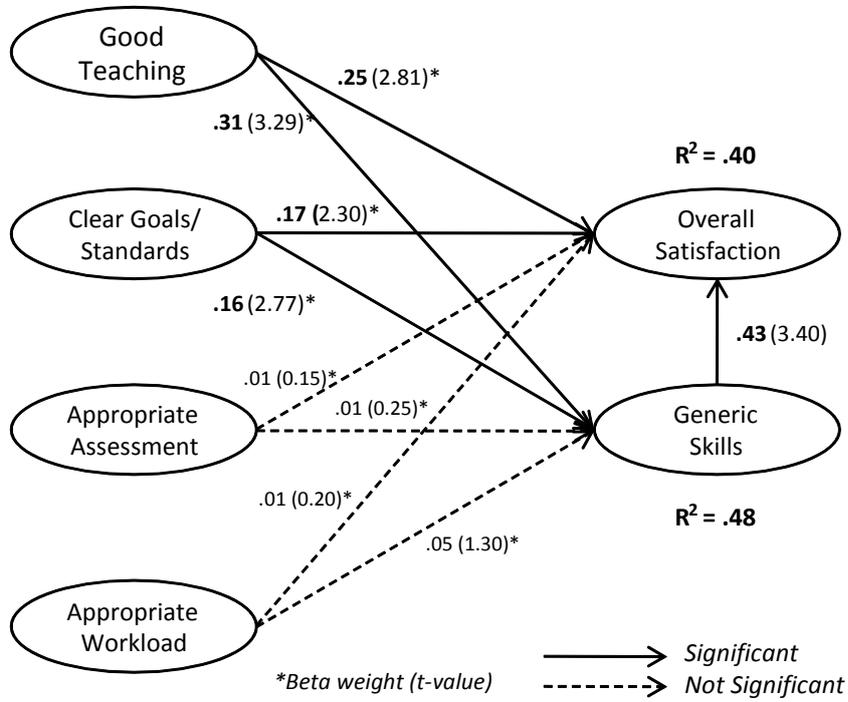


Figure 2: Model Results - Gender Comparisons

