RESEARCH ARTICLE

University performance evaluations: What are we really measuring?

Rowan E. Bedggood* **and Jerome D. Donovan**

Marketing and International Studies Group, Swinburne University of Technology, Melbourne, Australia

Abstract

Surveying students to garner feedback on teaching and subject quality are common occurrences in many universities globally. Despite the criticisms surrounding whether measures associated with these surveys are indeed valid, university managers continue to utilise them in key decision making. These surveys mirror business practices where measuring customer satisfaction via surveys is common. However, some argue that universities are misdirected in measuring satisfaction as a proxy for teaching quality, possibly subverting the potentially conflicting objective of student learning. Even so, both student satisfaction and student learning can be relevant performance measures. Accordingly, we have developed two robust measures of these constructs. We argue that student learning can be measured and used to provide formative feedback for improving teaching effectiveness. Alternatively, student satisfaction can be appropriate for determining whether students are "enjoying" their studies, and likewise offer distinct benefits to university managers measuring performance outcomes.

Keywords: student surveys; satisfaction; learning; teaching evaluations.

* Corresponding author. Email: RBedggood@swin.edu.au

Introduction

Universities have been increasingly confronted by shifting operational realities brought about by growing levels of international and domestic competition, changing demographics and shifting pressures from community and internal stakeholders (Driscoll & Wicks, 1998; Pike, 2004). In an attempt to attract and retain students, many universities have been lured into adopting a 'business-like' paradigm, and in particular, the measuring of "customer" (student) satisfaction (Sohail & Saeed, 2003; Lutz & Field, 1998). In so doing, they effectively mirror the private sector's marketing practice by regularly measuring "customer" satisfaction. While some may find 'students as customers' potentially distracting (*e.g.,* Maguad, 2007; Sharrock, 2000; Driscoll & Wicks, 1998; Sirvanci, 1996), others have suggested that these student satisfaction surveys are also purported to evaluate either the quality of teaching or the subject being delivered (Bedggood & Pollard, 2001; Craig, 1995). This has created some confusion regarding the ultimate purpose of student surveys; are they meant to be capturing teaching quality or student satisfaction? In response to this dilemma, this study will develop distinct measures of student learning (as a key outcome of teaching quality and a pillar driving the higher education system) and student satisfaction (emerging from the operational and competitive dynamics now facing universities), and thus present a way forward in measuring student experiences at university.

Student Surveys

Measuring Teaching Quality or Student Satisfaction?

Customer satisfaction remains an important aspect for businesses and a construct that is regularly measured in the private sector. Given the current competitive climate in which universities operate, it is unsurprising that they would focus on students and, like commercial businesses, regularly measure (customer) satisfaction. However, these surveys are described as "teaching evaluation" surveys, and interestingly, many scholars argue that these student surveys are measures of student "satisfaction" rather than measures of teaching quality (*e.g.,* Bedggood & Pollard, 2001; Craig, 1995; Harnash-Glezer & Meyer, 1991; Kleiner, 1989). Others go further, arguing that student surveys capture little more than the match between personalities and learning styles of students and their teacher and that they carry the added danger that "students may give the highest satisfaction ratings to those teachers who challenge them the least" (*e.g.,* Brookfield, 1996: 62).

Lending credence to Brookfield's (1996) assertion are the *Dr Fox* studies carried out in the 1980s, where it was shown that teacher expressiveness can have an overriding influence on student evaluations of teaching (student ratings) (Marsh, 1987). Ware and Williams (1980) reviewed several similar studies and conclude that differences in teacher expressiveness consistently explain more variance in student ratings than differences in content. A decade later, Atamian and Ganguli (1993) found that the majority of students surveyed (76%) perceive their least favourite instructor to also be the least effective, describing "favourite" and "effective" as "enthusiastic" and "energetic" respectively. These findings imply that high student ratings may be more a result of the popularity of the teacher than from "teaching quality" or indicators of students having learned (Ramdsen, 1990). This conclusion is supported by Clayson's (1999) review of previous research in the area, where he reports that between 50-80% of the variance in teacher evaluation surveys can be attributed to aspects related to instructor personality. Indeed, Guolla (1999) found instructor

enthusiasm to have the greatest impact on student "satisfaction" with instructors. It would appear that the *Dr Fox Effect* is still thriving.

To empirically test whether student satisfaction surveys are masquerading as teaching evaluation surveys, Bedggood and Pollard (2001) examined the content of over 800 survey items found in 30 "teaching evaluation surveys" used across eight Australian universities. They found that the surveys are predominantly measures of student satisfaction as opposed to measures reflecting aspects relevant to student learning in higher education (a more apt indicator of teaching quality). Consistent with these findings is the research conducted by Wiers-Jenssen, Stensaker and Grogaard (2002), who found an overlap between student satisfaction and teaching evaluation surveys. This becomes alarming when considering that Bedggood and Pollard (1999) found that these surveys are misused tools that lack reliability and validity. Given that few alternative or complementary methods of evaluation are used by university managers to evaluate staff (Ramsden, 1990), then by implication, academics are evaluated based partly on student satisfaction, using surveys that may well produce invalid and unreliable data. As it stands, the use of student satisfaction to evaluate teaching quality is unjustifiable, yet this is common practice (Bedggood & Pollard, 1999).

Despite these concerns, and perhaps due to a lack of easy-to-use alternatives in evaluating academic staff, university managers currently rely on student survey data. Garnering student feedback may be useful, though danger resides in how the data are interpreted and applied. For example, using student feedback to make strategic changes to subjects (or its delivery), or tenure and promotion decisions, could lead to erroneous decision making, if the data are flawed (Ramsden, 1990; Apodaca & Grad, 2005), yet this is also current practice. Further problems emerge when low evaluation scores are ascribed to poor teaching quality, when other salient reasons, such as the personality of the student/instructor, situational factors (economic or environmental context), and student motivation, could have produced low scores (Simpson & Siguaw, 2000). It is thus vital to validate and contextualise the data before relying on it for managerial decision making. We now examine the factors that require consideration to ensure the data is contextualised, or qualified.

Qualifying Student Satisfaction Data

Many factors associated with student satisfaction could, if not controlled for, contaminate or bias the findings. For example, McInnis *et al*., (1995) found that country of birth, gender and age of the student were significantly correlated with student satisfaction. Thus, students who were male, under 20 years of age, and/or born in South East Asia were found to be generally *less* satisfied with their course and teaching than other students. They also found that students from independent private schools, living with their family, or with highly educated parents, are likely to be *less* satisfied than other students. Likewise, Hendry (1983) found that age, gender and previous education level affect satisfaction, with females, older students, and students with lower educational levels, generally more satisfied than their counterparts. Class size has also been found to contaminate results, with smaller classes typically giving higher ratings (*e.g.,* Greenwald & Gillmore, 1997). Environmental and personal factors may also influence satisfaction. For example, in a national survey of 10,000 participants, Wiers-Jenssen *et al.,* (2002) found that social climate and aesthetic aspects of the institute account for approximately 29 per cent of the variance in student satisfaction. Also, Chow (2005) found a positive relationship between life satisfaction and certain aspects

of student satisfaction (academic experience and self-esteem). It would appear, then, notwithstanding students' demographic traits, if they are dissatisfied with aspects of their life (home, marriage, work commitments, or financial situation), then corresponding satisfaction levels at university are likely to degrade. These factors are mostly beyond the control of the university, and all factors cited here are beyond the control of the instructor, the latter being the recipient of student "ratings".

Auxiliary services have been identified as factors that can also influence student satisfaction levels. Stewart (1991) found the areas in which students were dissatisfied to include parking facilities, course scheduling procedures, student convenience, innovative tuition payment methods and quality of administrative staff services. The effect of some of these factors on student satisfaction has been verified in other studies (*e.g.*, Hendry, 1983; Liegler, 1997; Wiers-Jenssen *et al.*, 2002). Feedback on such factors may help guide maintenance and administrative decisions, but shed little light on the learning experience of students, nor the performance of academic staff.

This means that current student surveys may not adequately capture teaching "quality" because they lack items related to student learning, and may only reflect teacher popularity or other extraneous factors. It also means that they provide a poor indication of student satisfaction, unless (at the least) extraneous variables are included in surveys to permit controlling for their bias in data interpretation. Given the institutional persistence of using these measurement tools, value therefore resonates with developing a clear understanding of what information *should* be captured by student surveys, and also in generating distinct measures that reflect aspects of critical importance to inform university managers and academics alike. The purpose of this paper is thus to present two distinct and purposefully developed measures of student "satisfaction" and student "learning", and so provide new tools for university managers for evaluating teaching and subjects. Each instrument can be used depending on whether managers want to know if students are learning, or whether students are happy (satisfied) during the process of learning.

Scale Development

Student Learning Constructs

Few would argue that the main outcome of education is for students to have learned. A fundamental purpose of a university is to embrace this concept and to facilitate learning at its highest level (Candy, Crebert, & O'Leary, 1994). Perhaps the most significant difference between universities and other educational institutions is the goal of moving students from a dependent based way of learning, towards independence or autonomy. This typically involves transformation, where students actively negate old structures (Mezirow, 1989; Stanton, 1981) and embrace new paradigms of thinking, feeling and behaving (Brookfield, 1996; Higgins, Hartley & Skelton, 2002). Ideally, as a result of the learning process, students should become increasingly capable of self-directed study and responsible for their own learning, while their reliance on instructors should gradually diminish (Boud, 1981; Brookfield, 1989; Mezirow, 1989; Pike, 1991).

Educational experts rarely consider this transition process to be "comfortable" or "enjoyable" for students because it often involves discomforting discrepancies between actual experience and current beliefs (Brookfield, 1989). McInnis, *et al*., (1995) suggest that

university education "ought not be cognitively comfortable for students" (p10) and "[who] need a good dose of benign neglect" (p36). Students typically respond with resistance and reticence to the challenge of developing autonomy (Boud, 1981; Candy, *et al*., 1994; Mezirow, 1989) which suggests they may find university education to be a difficult task that is not always "satisfying", particularly during the learning process. Given the orientation of universities towards achieving the fundamental goal of student learning, it is surprising that this would be the area where macro-measurement indicators are least refined (Ramsden, 1991).

Redressing this shortfall are some studies which sought to articulate the relationship between students' perceptions of teaching and the quality of student learning (*e.g.,* Ramsden & Entwistle, 1981; Meyer & Muller, 1990; Ramsden, 1991; Espeland & Indrehus, 2003; Richardson, 2005). The investigation initiated by Ramsden and Entwistle (1981) prompted many studies, including one by the Australian federal government to study university performance indicators. Initially this Course Experience Questionnaire included five broad scales including good teaching, clear goals and standards, emphasis on independence, appropriate assessment and appropriate workload (Richardson, 2005). Adaptations to this model have been progressively offered, although it was the addition by some, such as McInnis, *et al*., (2001) that increased the scope of measurement towards learning with the inclusion of five additional scales; including student support, intellectual motivation, learning community, graduate qualities and learning resources scales.

While the outcomes of learning can be judged from the perspective of the university, instructor and student (*e.g.,* Entwistle & Smith, 2002; Hay, 2007), instead of measuring "learning" from a teaching perspective, as many have done before (*e.g.,* Adpodaca & Grad, 2005; Ginns, Prosser & Barrie, 2007; Liegler, 1997), we adopt a student perspective to measuring their learning within the classroom. Retaining this student-centric focus is shared by others (*e.g.,* Shipengrover and James, 1999; Espeland & Indrehus, 2003). We propose that adopting this perspective will overcome many of the shortfalls of relying on "broader" student feedback as a pseudonym for evaluating teaching quality. From this student perspective, learning outcomes can be measured using test/assignment/exam results, final grades, grade-point-average (GPA) across the whole course, the number of subjects passed, or a combination of all (Docking & Thornton, 1979; Sarid, Anson, Yaari & Margaltih, 2004). While grades appear to be an objective measure of learning, they can mask subtle nuances during the learning process, for example, the extent to which students are learning, and whether they are doing so at the required level. It could be that student grades reflect prior knowledge, rather than current learning. Grades can also reflect intellectual capability, which may subtly differ from whether students "perceive" they have learned. Consequently, this study focuses on students' perceptions of learning in higher education. With this in mind, we extend the platform offered by Entwistle and Smith (2002), arguing that the key dimensions within student learning can be summarised as:

1. ***Increases in knowledge and skills*** (perceived learning)
   This type of learning can be conceptualised as the abstraction of meaning; an interpretive process aimed at understanding reality (Kember, Sivan & Davies, 1995), generating alternative ways of thinking (Brookfield, 1989), the personal discovery of meaning (Stanton, 1981), critical thinking (Pike, 1991) and transformation (Mezirow,

1989). Students' perception of learning will provide greater insight thus overcoming the shortfalls of using grades alone.

2. ***Stimulation and motivation*** (perceived challenge)
   The degree to which students feel challenged aims to capture the higher learning notions presented by Brookfield (1996), Candy, *et al.*, (1994) and Entwistle and Smith (2002) wherein students are both motivated and stimulated.
3. ***Complexity and difficulty*** of concepts (perceived task difficulty)
   Student perception of whether a task is difficult has been found to affect learning (Espeland & Indrehus, 2003) while also capturing the notion that higher educational learning ought not to be cognitively comfortable for students (McInnis, *et al*, 1995).
4. ***Self-evaluating performance*** (perceived performance).
   Student perception of their performance aims to capture the self-reflective nature desired in higher educational learning (Boud, 1981).

A paucity of pre-existing higher educational learning scales testifies to the difficulty of transposing these concepts into a measurement instrument. To fill this gap, we generated a battery of items (16) for these four dimensions in preparation for testing and refining.

Student Satisfaction Constructs

Unlike student "learning" scales, an abundance of studies have produced scales in an attempt to measure student "satisfaction". An overview of previous studies reveals that there are key dimensions commonly used to capture student satisfaction within a single subject. It would appear that aspects pertaining to the "quality of instruction" typically account for most of the variance and, as such, dominate the satisfaction response. For example, Harnash-Glezer and Meyer (1991) identified six factors of student satisfaction in the class, however, "instructor ability and knowledge" was the most significant contributor to student satisfaction (representing 46% of the variance), and well above the other factors which each contributed less than 5% of the variance. Similarly, Liegler (1997) found that "instructor knowledge and expertise" (together with one other construct) accounted for 68% of the variance in student satisfaction. Several others have also used instructor related items to capture student satisfaction (*e.g.*, Chadwick & Ward, 1987; Betz, Klingensmith, & Menne, 1970). It appears that this dimension is a critical one to represent in a student satisfaction survey.

From its origins, the satisfaction component of "quality of instruction" has since been extended. For example, Hendry (1983) developed a 44-item survey measuring four factors of student satisfaction. Psychometric testing revealed two separate dimensions of satisfaction with teaching: 1) "interpersonal relations" (friendliness of, and interactions with, staff, faculty and other students), and 2) "program quality" (instructor skills and knowledge, job preparation and grading fairness). This was also the decade of the *Dr Fox* studies, discussed earlier, where instructor enthusiasm and skills were found to have opposing effects on student learning and satisfaction, thus providing empirical fuel to the debate over whether students were truly capable of evaluating teaching effectiveness or whether their evaluations were reflections of how much they "liked" the instructor. Interestingly, subsequent studies on the dimensionality of student satisfaction rarely continued the separation between "skills" and "friendliness" (*e.g.*, Harnash-Glezer & Meyer, 1991; Krehbiel, McClure and Pratsini, 1997) and instead, teacher enthusiasm and knowledge are typically clumped under the single construct of "instructor ability and knowledge".

Several other dimensions have been identified over the decades. From the 1970s the concept "effort required" by the student also emerged. This refers to the effort required by students to achieve their desired grades (*e.g.*, Betz, *et al.*, 1970). This second construct has been operationalised in a variety of ways. For example, Harnash-Glezer & Meyer (1991) identified "easiness of demands" and included one item on student effort invested in the course as a dimension of student satisfaction, while Krehbiel, *et al*., (1997) identified both "content difficulty" and "grading fairness" as satisfaction dimensions. If it is assumed that difficult tasks require greater effort from the student, then it could be argued that "task demands" form a component of student learning, as well as a component of student satisfaction. In combining the ideal that university education ought to challenge students, and that difficult tasks may diminish satisfaction, then by implication, the difficulty of a subject must be high enough to engage and challenge students, but not too high or their satisfaction will be eroded.

A final construct has emerged more recently: some studies have identified "academic development and stimulation" as a valid construct of student satisfaction (*e.g.*, Harnash-Glezer & Meyer, 1991; Liegler, 1997). This refers to how satisfied students feel with the process of learning; whether they feel stimulated by their experience and whether they feel they are growing and developing in their academic ability. Similar to "task difficulty", this construct is also in keeping with the broader epistemological views often espoused by higher educational experts (Boud, 1981; Mezirow, 1989; Ramsden, 1990; Brookfield, 1996; Candy, *et al*., 1994; McInnis, *et al*., 1995).

This analysis of satisfaction surveys therefore reveals three commonly identified dimensions of student satisfaction:

1. ***Quality of Instruction*** refers to both instructor skills and ability, and also to their friendliness, enthusiasm and approachability.
2. ***Task Difficulty*** in terms of demands and effort required by students to achieve their desired result.
3. ***Academic Development and Stimulation*** refers to how stimulated and motivated a student feels and whether they believe they are growing and developing their academic skills.

This review reveals that no single existing measure captures each of these constructs with sufficient depth and in the one instrument. Moreover, many Faculties and the subjects offered therein have parochial aspects pursuant only to the respective discipline. Accordingly, many previous measures capturing student satisfaction have been developed with a specific discipline in mind (*e.g.*, Liegler, 1997; Atamian & Ganguli, 1993). Thus, few efforts have been made to develop surveys suitable for measuring satisfaction within the classroom, and which can be generalised across most disciplines. This study seeks to rectify this shortfall. Using items drawn from previous studies, relevant ones were extracted, some wording was modified and new items developed, which produced a final 43 items to capture student satisfaction.

Methodology

The procedures recommended by Churchill (1979) and Peter (1979) have been used to guide the development, testing and refinement of each scale and accordingly, a three-stage process was adopted to ensure the final survey instrument was as rigorous as possible. First, a pilot study was conducted on a small cohort of students in a second year business subject. This stage was vital for it allowed exploratory factor analysis (EFA) to reduce the large number of initial items. Once this survey was refined, it was then administered to a new cohort of students, in class, in two second year business subjects across two campuses of a large Australian university. After further refinement using EFA, the survey was then administered to a larger cohort of students across the same two subjects and campuses. This final data set was treated to Structural Equation Modelling (SEM) to establish the confirmatory factor analytical (CFA) structure of both measures. The psychometric properties of the instruments were further verified which produced two robust measures of student satisfaction and learning. The results of the second EFA and final CFA are presented here.

All surveys were personally administered during class, which may account for the high response rate achieved, producing 81% overall, with 86% in stage one (n=76), 89% in stage two (n=341) and 78% in stage three (n=874). The respondent profile for those participating in steps two and three (94% of all respondents) was: 56% female, 96% completing a Bachelor of Business and 50% were 20-22 years of age. Most students were either permanent residents of Australia (67%) though only 48% reported Australia as their country of origin, with 37% originating from Asian countries. Most students were "single" (81%) and had been at university for more than one year (75%), thus the university learning experience for most participants was no longer novel.

Results

EFA was conducted using principal component factor analysis to identify the underlying dimensions of the two scales for the first two stages. To determine which items were acceptable in the factor structure, the procedure and evaluation points presented by Hair, Anderson, Tatham and Black (1998) were followed. At each stage, if a value associated with an item fell below the "minimum" range suggested by these authors, then it was removed, and the analysis re-conducted. This iterative process provided a rigorous method of item reduction while also enhancing the robustness of each measure. Stage one, using an oblique rotation, produced the expected four factors for learning, and four factors for satisfaction: the expected three, though the "first" split into two, revealing that students make a clear distinction between the skills of an instructor and the personable aspects of an instructor.

Main exploratory study

EFA was again conducted on data from a new and larger cohort of students. However, EFA can sometimes exclude items or aspects of a construct leaving a "bare-bones" measure of the factor. Following Hair, *et al*., (1998) in their recommendations for scale development, after the initial item-reduction process (first EFA), it is prudent to reintroduce some items after modifying the wording of the question, or to develop a new question that somehow captures several that were previously excluded. This process provides the opportunity to re-

ingest some "flesh" into each factor, thus ensuring that critical aspects are not omitted from the final survey. Tables 1 and 2 indicate where such adjustments were made to each of the measures and present the factor solutions for each dimension, together with reliability and validity test results.

Student Learning: EFA on stage two was conducted using varimax rotation (due to a lack of correlation among the constructs) and, after deletion and reintroduction of some items, produced a 12-item three-factor solution: task difficulty and challenge now loaded on the same factor. Cronbach's alpha reveals that the challenge factor has acceptable reliability, and the others very good reliability (see Table 1). For this EFA, the items previously loading under "challenge" loaded instead with the "perception of learning" factor. Students from this larger cohort did not perceive that being intellectually stimulated and motivated to achieve their best are separate from developing their understanding, skills and general learning. With another "challenge" item reinstated for this stage, both items then loaded with perceptions of difficulty. The results from this analysis suggest that students do not discern between the concept of challenge and task difficulty. Accordingly, these items were then grouped under the term "challenge".

PLEASE INSERT TABLE 1 HERE

Student Satisfaction: The EFA solution for student satisfaction retained a total of 19 items. However, several changes were made to the scale between the pre-test stage and main study stage. Some items were reinstated from the original survey, and one new item was generated. The item "degree of challenge" changed from loading under the course difficulty/equity factor in the first EFA, and instead loaded under the academic development and stimulation factor for the larger cohort in the second EFA. This suggests that students may view "challenge" as a positive contribution towards their academic development, rather than as a sense of "difficulty" during the learning process. Interestingly, satisfaction with personable aspects of the instructor was again the first factor extracted. All Cronbach alpha's attest to the persistent reliability of each of the four factors (see Table 2).

PLEASE INSERT TABLE 2 HERE

Confirmatory study

Confirmatory Factor Analysis using SEM on the data from a new cohort of students was the final step in scale purification. SEM is essentially confirmatory, as it determines the extent to which the proposed structure is consistent with the empirical data (Diamantopoulos, 1994). A range of goodness-of-fit indices were used to indicate how well the structural model fit the data. Since no single measure adequately captures the strength of the model's predictions (Hoyle & Panter, 1995), a number of indices were used in combination to indicate acceptable levels of fit, based on: i) overall fit, ii) comparative fit to a base model and iii) model parsimony. The guide provided by Schumacher and Lomax (1996) on acceptable levels for each of the indices were used to inform this analysis.

Student learning: To confirm the construct validity for the student learning scale, the 12-items were tested using *AMOS 4*. The results indicate that the model has acceptable fit with the data and supports a 3-factor solution for student learning (see Figure 1). The goodness-of-

fit indices are all healthy at over 0.90, particularly the TLI, NFI and CFI which were each over 0.99 and the RMSEA at 0.057 represents a very good fit. The Cmin/df is just higher than three and the pclose is somewhat low at 0.113, although, due to the model's complexity, this still indicates a good fit between the model and data, indicating reasonably good construct validity.

PLEASE INSERT FIGURE 1 HERE

Further tests of validity were conducted during this stage. Discriminant validity is evident by the low correlations between each of the constructs, which means that each construct captures a unique latent variable. To determine convergent validity, a global item *"how would you rate your academic experience in this subject*" was correlated with perceived learning. The correlation between the two is 0.373 ($p<0.01$). Although this is an acceptable fit, there is room to improve the degree of convergent validity of the perceived learning scale. It could be argued, for example, that this global question spans more than "perceived learning" thus explaining the somewhat low correlation.

In order to establish criterion validity of the perceived performance construct, students' grades during semester were matched with their corresponding survey responses on perceived learning and performance. It is expected that if students believe they are performing well, then their grades will be higher than those students who do not. Using a Pearson Moment correlation, it was found that the coefficient for the relationship between perceived learning and final grade was 0.175 ($p<0.01$). The correlation between perceived performance and final grade was identical. Subsequent regression analysis shows that perceived learning and perceived performance explain a meagre 5% of the variation in final grade. Although the validity of these new measures is acceptable, it also demonstrates there may be room to improve the robustness of these factors. Alternatively, these results may simply reflect that grades are not an "absolute" when it comes to measuring whether students think they are learning and performing well (as previously discussed).

Student satisfaction: A measurement model for student satisfaction was constructed that supported the four-factor solution from the EFA in stages one and two (see Figure 2). The results obtained from the SEM indicate that the model has acceptable fit with the data. The Cmin/df is between one and three, which is good for a complex model. The other goodness-of-fit indices indicate a good fit between the model and data, while the RMSEA is also acceptable at less than 0.05. The AIC figure is quite high. However compared to the independence model (11131.267) it is acceptable. Overall, the results of the SEM indicate sound construct validity of this measure of student satisfaction.

PLEASE INSERT FIGURE 2 HERE

Some items were included in the survey specifically to test for the discriminant validity of the satisfaction measure. By including items that address areas of satisfaction not pertaining to a single subject, it was ensured that the measure is not an artefact of the lack of other factors to explain the variance in student satisfaction. For example, student satisfaction with "life in general" and the "university in general" (which have been used elsewhere, *e.g.*, Chow, 2005; Rain, Lane & Steinder, 1991; Witt & Handal, 1984), served as the discriminant

measures of student satisfaction in the present study. Table 3 presents the correlations between these broad satisfaction constructs and the four dimensions of student satisfaction in the classroom. These two items were not highly correlated with the components of satisfaction for a subject, indicating sound discriminant validity of the satisfaction measure.

PLEASE INSERT TABLE 3 HERE

To test for convergent validity, overall satisfaction measures were included in the surveys for each of the four sub-constructs. Using Pearson Moment correlations, significant and strong correlations were found between the respective global item and: satisfaction with instructor skills (0.61, $p<0.01$); satisfaction with instructor personable aspects (0.63, $p<0.01$); satisfaction with the course difficulty/equity (0.50, $p<0.01$); satisfaction with academic development/stimulation (0.41, $p<0.01$). Hence, the first three constructs show good convergent validity, while the last construct shows moderate convergent validity. The final test for convergence was to compare the summated satisfaction score with the overall satisfaction with subject quality score. The correlation was significant between the global and summated scores (0.55, $p<0.01$). In summary, the psychometric tests indicate that the newly developed measures of student satisfaction and student learning are robust and indicate sound validity and reliability.

Discussion/Conclusion

With the adoption of marketing ideals such as "customer satisfaction" we have witnessed the pervasive use of student surveys by university administrators as a way of measuring "teaching quality", reflecting a customer-centric focus on students. This study has sought to refocus the ongoing academic debate about the risks associated with proclaiming students as customers (e.g., Sharrock, 2000; Driscoll & Wicks, 1998) or using student survey data *per se*, to the measures themselves. We recognise that university managers will continue the practice of regularly collecting feedback from students, and so there remains a greater need for clarity in what is being measured by these surveys. Accordingly, we developed two robust measures to reflect the two, potentially competing, objectives of university education: the traditional objective of *student learning* and the more recently emergent objective of *student satisfaction*.

The results of this study confirm that universities do indeed need two unique measures to gauge the extent that subjects are fulfilling these dual objectives. It is quite clear from existing literature that current surveys for measuring teaching quality, educational quality, and for benchmarking broader organisational performance, have been focused on student satisfaction rather than student learning (*e.g.*, Bedggood & Pollard, 2001; Wiers-Jenssen, *et al.*, 2002). This study reaffirms the need to treat these two objectives of universities differently, and to therefore use two distinct scales to measure each.

The findings presented here illuminate the unique nature of both constructs. Specifically, this study has demonstrated empirical support for theorists regarding key facets of student learning, demonstrating that a survey should comprise items reflecting: student knowledge and skills; the extent to which they are stimulated, motivated and challenged; and their self-evaluating performance. This student-centric approach should resonate with educationalists who espouse that the effectiveness of courses and instructors should be viewed in terms of

student performance or learning outcomes (March, 1987; Pike, 1991; Ramsden & Dodds, 1989).

This study has also developed a robust survey for capturing student satisfaction. The results support the contention that when measuring student satisfaction, both the personable and skills aspects of the instructor should be measured, and reported, separately. Furthermore, student satisfaction with non-teaching aspects of a subject, such as academic development and course difficulty/equity are vital to include if student satisfaction is to be fully captured. It is also suggested that, given the numerous extraneous factors that can affect the student experience, university survey administrators should also make every effort to measure at least some of these factors in their surveys (*e.g.*, student demographics, life satisfaction, class-size, *etc.*) to reduce bias and contextualise the results accordingly. We also suggest that summary data per sub-construct be included in reports. Moreover, this study supports the views of Ramdsen (1991) and Apodaca and Grad (2005), recommending that administrators do not rely solely on student surveys to evaluate "teaching quality". Where this is not possible, it is recommended that the student learning scale developed here is used in preference to student satisfaction-type measures (or "teaching evaluation" surveys, as they are sometimes termed) (Bedggood & Pollard, 2001), because it carries greater criterion validity of "teaching quality" than student satisfaction.

The findings of this study also identify that student learning and satisfaction share two common dimensions: *academic development/stimulation* and *task difficulty/challenge*. It could be argued that current surveys, therefore, already capture two of the suggested constructs developed from our study. However, it is the approach of each that reveals their difference: the learning component captures the process or occurrence of learning (*I feel stimulated* reveals the extent of the stimulation), while the satisfaction component captures the emotional and cognitive judgement (or evaluation) of the occurrence of having learned (*I like how much I am stimulated* reveals enjoyment from being stimulated). It seems intuitive that these two constructs would indeed be interrelated, suggesting that academic teaching staff need to be aware of the difference, and be careful in creating a balance between the level of challenge (and potential to push students' capabilities) and the level of satisfaction students feel about being challenged. It is possible that the *dumbing-down* of standards is not the "slippery slope" that some have feared (*e.g.*, Clarke, 1998; Driscoll & Wicks, 1998).

There is likely to be a strong link between the extent to which students are engaged with their university experience, and their degree of satisfaction with being thus engaged. A more engaged student is likely to also optimise the extent of their learning. Indeed, the relationship between student satisfaction and learning has been investigated by others, but with unclear and often contradictory findings. For example, Pike (1993) comments that his previous empirical findings (Pike, 1991) demonstrate that the relationship between grades and satisfaction is weak. Guolla (1999) found that student learning was more closely associated with course satisfaction than instructor satisfaction, while Donahue and Wong (1991) found a positive correlation between student achievement and only one of five satisfaction components. It would be useful if future research further investigated the extent of the relationship between these prime variables.

While we argue against using student satisfaction as a pseudo measure of teaching quality, it remains an invaluable dimension. Understanding student satisfaction will illuminate the depth of the total student experience, because it draws upon the combined effect of students' *cognitive* and *emotional* evaluation. This is well recognised as being meaningful by marketers in the commercial sector measuring "customer" satisfaction (e.g., Oliver, 1993). Furthermore, understanding student satisfaction holds some credibility, because it relates to areas in which educators place great importance, notably academic achievement (Pike, 1993) and attrition (Koseke and Koseke, 1991). Moreover, it should be noted that in the educational setting, the role that functional aspects of marketing can play is not usually challenged, such as determining appropriate marketing strategies (Licata & Frankwick, 1996), promotional strategies and how to export educational services (Mazzarol & Hosie, 1996). Accordingly, it is argued that measuring and responding to student satisfaction could produce favourable outcomes for universities and students alike, as systems and process can be adjusted to make the university experience more enjoyable for students, which may influence favourable word-of-mouth and enhance university image and reputation.

The findings presented here should be interpreted in light of some limitations. Although the satisfaction and learning scales display sound psychometric properties, neither scale accounted for more than 60-70% of the variance of the construct in the second EFA. Also, further scale improvement would enhance the good reliability levels established here. This means that there is scope for future research efforts to continue refining these instruments. Although we drew from non-discipline specific views to develop items for these scales, most students in each sample were studying a business degree, which could contain inherent bias. This leaves scope for testing the robustness of these measures with students studying other disciplines.

On a final note, managers are cautioned against blithely using both measures developed here and thus unnecessarily perpetuating the "evaluation machine" by over-surveying students. Rather, our surveys could be used to capture student learning, as one indicator of teaching quality, every two years, and to capture student satisfaction every 2-3 years. Meaningful, regular changes from one semester to the next are unlikely and thus surveying students this frequently risks wasting the time for everyone involved.

References

Apodaca, P. & Grad, H. (2005) The Dimensionality of Student Ratings of Teaching: Integration of Uni- and Multidimensional Models, *Studies in Higher Education*, 30(6), 723-748.

Atamian, R. & Ganguli, G. (1993) Teacher Popularity and Teaching Effectiveness: Viewpoint of Accounting Students, *Journal of Education for Business*, 68(3), 163-69.

Bedggood, R. E. & Pollard, P. (1999) The Uses and Misuses of Student Opinion Surveys in Eight Australian Universities, *Australian Journal of Education*, 43(2), 129-41.

Bedggood, R. E. & Pollard, R. (2001) Customer Satisfaction and Teacher Evaluation in Higher Education, in *Proceedings from the World Marketing Congress,.0* Cardiff, Wales.

Betz, E. L., Klingensmith, J. E. & Menne, J. W. (1970) The Measurement and Analysis of College Student Satisfaction, *Measurement and Evaluation in Guidance*, 3(2), 110-18.

Boud, D. (1981) *Toward Student Responsibility for Learning*, in Developing Student Autonomy in Learning, David Boud, Ed. New York: Kogan Page, Nichols Publishing Company.

Brookfield, S. (1989) *Teacher Roles and Teaching Styles*, in Lifelong Education for Adults: An International Handbook, Colin A. Titmus, Ed. Oxford, England: Pergamon Press.

Brookfield, S. (1996) Experiential Pedagogy: Grounding Teaching in Students' Learning," *Journal of Experiential Education*, 19(2), 62-68.

Candy, P. C., Crebert, G & O'Leary, J. (1994) *Developing Lifelong Learners through Undergraduate Education*, Melbourne: Australian Government Publishing Service, National Board of Employment, Education and Training.

Chadwick, K. & Ward, J. (1987) Determinants of Consumer Satisfaction with Education: Implications for College & University Administrators, *College & University*, 62, 236-46.

Chow, H. P. H. (2005) Life Satisfaction Among University Students in a Canadian Prairie City: A Multivariate Analysis, *Social Indicators Research*, 70(2), 139-50.

Churchill, G. A. Jr. (1979) A Paradigm for Developing Better Measures of Marketing Constructs" *Journal of Marketing Research*, 16 (February), 64-73.

Clarke, H. (1998) Dumbing-Down in Australian Universities, *Quadrant* (42), 55-59.

Craig, T. A. (1995) Student Satisfaction Measurement (SSM) in Ontario Universities as Perceived by Student Services Leaders and Others, *College Student Journal*, 29 (Dec), 430-37.

Diamantopoulos, A. (1994) Modelling with Lisrel: A Guide for the Uninitiated, *Journal of Marketing Management*, 10, 105-36.

Docking, R. A. & Thornton, J. A. (1979) Anxiety, Achievement and Cognitive Incongruity, *Australian Journal of Education*, 23(3), 250-61.

Driscoll, C. & Wicks, D. (1998) The Customer-driven Approach in Business Education: A possible Danger?, *Journal of Education for Business, 74(1),* 58-61.

Entwistle, N. & Smith, C. (2002) Personal understanding and target understanding: mapping influences on the outcome of learning, *British Journal of Education Psychology*, 72, 321-342.

Espeland, V. & Indrehus, O. (2003) Evaluation of students' satisfaction with nursing education in Norway, *Journal of Advanced Nursing*, 42(3), 226-236.

Ginns, P., Prosser, M. & Barrie, S. (2007) Students' perceptions of teaching quality in higher education: the perspective of currently enrolled students, *Studies in Higher Education*, 32(5), 603-615.

Guolla, M. (1999) Assessing the Teaching Quality to Student Satisfaction Relationship: Applied Customer Satisfaction Research in the Classroom, *Journal of Marketing Theory and Practice*, 7(3), 87-97.

Greenwald, A. G. (1997) "Validity Concerns and Usefulness of Student Ratings of Instruction", *American Psychologist*, 52, 1182-1186.

Greenwald, A. G. & Gillmore, G. M. (1997) Grading Leniency is a Removable Contaminant of Student Ratings, *American Psychologist*, 52, 1209-1217.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998) *Multivariate Data Analysis* (Fifth ed.), New Jersey: Prentice-Hall International, Inc.

Harnash-Glezer, M. & Meyer, J. (1991) Dimensions of Satisfaction with College Education, *Assessment & Evaluation in Higher Education*, 16(2), 95-107.

Hay, D. B. (2007) Using concept maps to measure deep, surface and non-learning outcomes, *Studies in Higher Education*, 32(1), 39-57.

Hendry, A. M. (1983) Measuring Adult Student Satisfaction: A Model, *Canadian Vocational Journal*, 19(1), 47-50.

Higgins, R., Hartley, P. & Skelton, A. (2002) The conscientious consumer: reconsidering the role of assessment feedback in student learning, *Studies in Higher Education*, 27(1), 53-64.

Hoyle, R. H. & Panter, A. T. (1995) *Writing about Structural Equation Models*. California: Sage Publications Inc., 158-176.

Kember, D., Sivan, A. & Davies, H. (1995) An Analysis of the Learning Process of Business Students at Hong Kong Polytechnic" *Journal of Education for Business*, 70(3), 172-179.

Koseke, G. F. & Koseke, R. D. (1991) Student "Burnout" as a Mediator of the Stress-Outcome Relationship, *Research in Higher Education*, 32, 415-31.

Krehbiel, T. C., McClure, R. H. & Pratsini, E. (1997) Using Student Disconfirmation as a Measure of Classroom Effectiveness, *Journal of Education for Business*, 72(4), 224-29.

Laurillard, D. (1993) *Rethinking University Teaching: A Framework for the Effective Use of Educational Technology,* London: Routledge.

Licata, J. & Frankwick, G. L. (1996) University Marketing: A Professional Service Organization Perspective, *Journal of Marketing for Higher Education*, 7(2), 1-16.

Liegler, R. M. (1997) Predicting Student Satisfaction in Baccalaureate Nursing Programs: Testing a Causal Model, *Journal of Nursing Education*, 36(8), 357-64.

Lutz, F. W. & Field, R. W. (1998) Business Valuing in Academia: The American University as a Center for Profit or Inquiry?, *Higher Education*, 36, 383-419.

Maguad, B.A. (2007), 'Identifying the needs of customers in higher education', *Education*, 127 (3), 332-343.

Marsh, H. W. (1987) Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research, *International Journal of Educational Research*, 11, 253-388.

Mazzarol, T. & Hosie, P. (1996) Exporting Australian Higher Education: Future Strategies in a Maturing Market, *Quality Assurance in Education*, 4(1), 37-50.

McInnis, C., James, R. & McNaught, C. (1995) *First Year on Campus. Diversity in the Initial Experiences of Australian Undergraduates,* Melbourne: Australian Government Publishing Service. Centre for the Study of Higher Education, University of Melbourne.

McInnis, C., Griffin, P. Richard, J. & Coates, H. (2001) *Development of the Course Experience Questionnaire (CEQ*), Centre for the Study of higher Education and Assessment Research Centre, Faculty of Education, University of Melbourne.

Meyer, J. H. F. and M.W. Muller (1990) Evaluating the Quality of Student Learning: An unfolding analysis of the association between perceptions of learning context and approaches to studying at an individual level, *Studies in Higher Education*, 15(2), 131-155.

Mezirow, J. (1989) *Personal Perspective Change Through Adult Learning,* in Lifelong Education for Adults: An International Handbook, Colin Titmus, Ed. Oxford, England: Pergamon Press.

Oliver, R. L. (1993) Cognitive, Affective, and Attribute Bases of the Satisfaction Response, *Journal of Consumer Research*, 20, 418-30.

Peter, P. J. (1979) Reliability: A Review of Psychometric Basics and Recent Marketing Practices" *Journal of Marketing Research*, 16 (February), 6-17.

Pike, G. R. (1991) The Effects of Background, Coursework, and Involvement on Students' Grades and Satisfaction, *Research in Higher Education*, 32, 15-30.

Pike, G. R. (1992) The Components of Construct Validity: A Comparison of Two Measures of General Education Outcomes, *Journal of General Education*, 41, 130-59.

Pike, G. R. (1993) The Relationship between Perceived Learning and Satisfaction with College: An Alternative View, *Research in Higher Education*, 34(1), 23-40.

Rain, J. S., Lane, I. M. & Steiner, D. D. (1991) A Current Look at the Job Satisfaction/Life Satisfaction Relationship: Review and Future Considerations, *Human Relations*, 44(3), 287-307.

Ramsden, P. (1990) Evaluating Teaching; Supporting Learning, in Research and Development in Higher Education, HERDSA Conference, Bob Ross (Ed.), Vol. 13. Griffith University, Brisbane: HERDSA.

Ramsden, P. (1991) A performance indicator of teaching quality in higher education: the course experience questionnaire", *Studies in Higher Education*, 16(2), 129-141.

Ramsden, P. & Dodds, A. (1989) *Improving Teaching and Courses. A Guide to Evaluation*, Parkville, Victoria: Centre for the Study of Higher Education, The University of Melbourne.

Ramsden, P. and Entwistle, N. J. (1981), 'Effects of academic departments on students' approaches to studying', *British Journal of Educational Psychology*, 51(3), 368-383.

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

Sarid, O., Anson, O., Yaari, A. & Margalith, M. (2004) Academic Stress, Immunological Reaction, and Academic Performance Among Students of Nursing and Physiotherapy" *Research in Nursing and Health*, 27(5), 370-77.

Sharrock, G. (2000) Why Students are not (Just) Customers (and Other Reflections on Life After George), *Journal of Higher Education Policy and Management*, 22(2), 149-64.

Shipengrover, J. A. & James, P. A. (1999) Measuring instructional quality in commsubjecty-orientated medical education: looking into the black box, *Medical Education*, 33, 846-853.

Simpson, P. M. & Siguaw, J. A. (2000) Student Evaluations of Teaching: An Exploratory Study of the Faculty Response, *Journal of Marketing Education*, 22(3), 199-213.

Sirvanci, M. (1996) Are Students the True Customers of Higher Education?," *Quality Progress*, 29(10), 99-102.

Sohail, M. S. & Saeed, M. (2003) Private Higher Education in Malaysia: Students' Satisfaction Levels and Strategic Implications, *Journal of Higher Education Policy and Management*, 25(2), 173-81.

Stanton, H. (1981) *Independent Study: A Matter of Confidence?*," in Developing Student Autonomy in Learning, David Boud, Ed. New York: Kogan Page, Nochols Publishing Company.

Stewart, K. L. (1991) Applying a Marketing Orientation to a Higher Education Setting, *Journal of Professional Services Marketing*, 7(2), 117-24.

Taylor, S. A. (1996) Consumer Satisfaction with Marketing Education: Extending Services Theory to Academic Practice" *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 9, 207-220.

Ware, J. E. & Williams, R. G. (1980) A Reanalysis of the Doctor Fox Experiments, *Instructional Evaluation*, 4, 15-18.

Wiers-Jenssen, J., Stensaker, B. & Grogaard, J. B. (2002) Student Satisfaction: Towards an Empirical Deconstruction of the Concept, *Quality in Higher Education*, 8(2), 183-95.

Witt, P. H. & Handal, P. J. (1984) Person-Environment Fit: Is Satisfaction Predicted by Congruency, Environment, or Personality?, *Journal of College Student Personnel*, 25(6), 503-08.

Worthington, A. G. & Wong, P. T. P. (1979) Effects of Earned and Assigned Grades on Student Evaluations of an Instructor, Journal of Educational Psychology, 71, 764-775.

Tables and Figures to insert

Table 1: Main Exploratory Factor Analysis for Perceived Learning

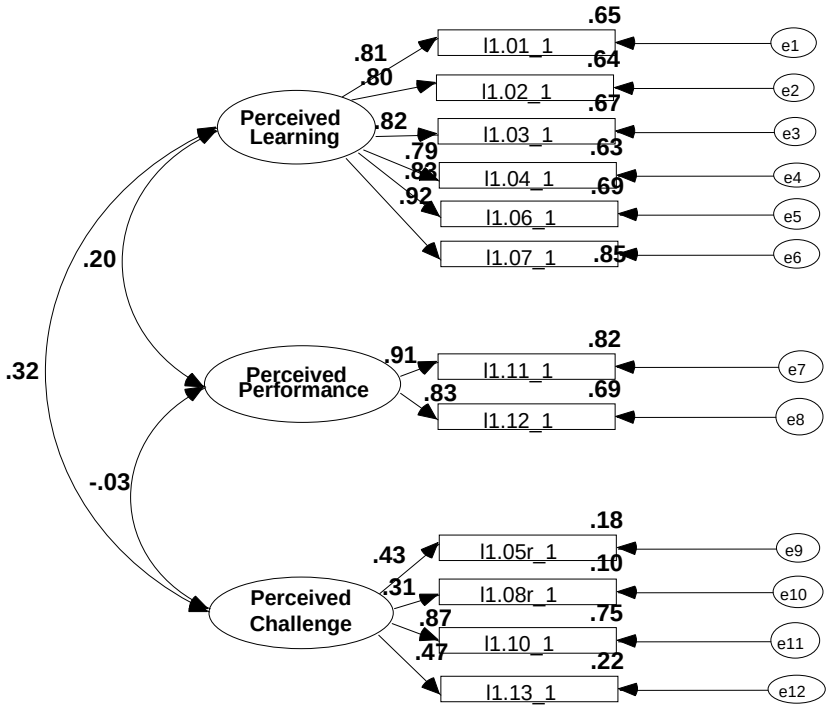
Learning / Achievement **n = 341** **Total Variance Explained = 62.9%**	**Factor 1 Perceived Learning**	**Factor 2 Perceived Challenge**	**Factor 3 Perceived Performance**
This subject is intellectually stimulating †	.833		
Developing skills in applying learned material	.804		
My interest in the topic has greatly increased †	.783		
I have been motivated to achieve my best †	.776		
So far in this subject I have learned a lot	.735		
Developing my understanding of key concepts	.713		
Assessment tasks have helped me to learn			
The subject demands are easy (R)		.773	
How would you rate the difficulty of this course		.708	
I have not yet been challenged (R)		.679	
I have found this subject to be challenging*		.536	
How does your performance compare with others			.868
How would you rate your performance			.845
What final grade do you expect to receive			
Eigenvalue	**4.148**	**2.095**	**1.306**
Percent of variance explained	**34.6**	**17.5**	**10.9**
Cronbach's alpha	**.8745**	**.6303**	**.8196**

* Denotes the item has been reinstated
** Denotes the item has been reworded
† Denotes the item is loading under a different factor
Items in italics indicate that they were discarded or not included in the main analysis

Table 2: Stage Two Exploratory Factor Analysis for Student Satisfaction

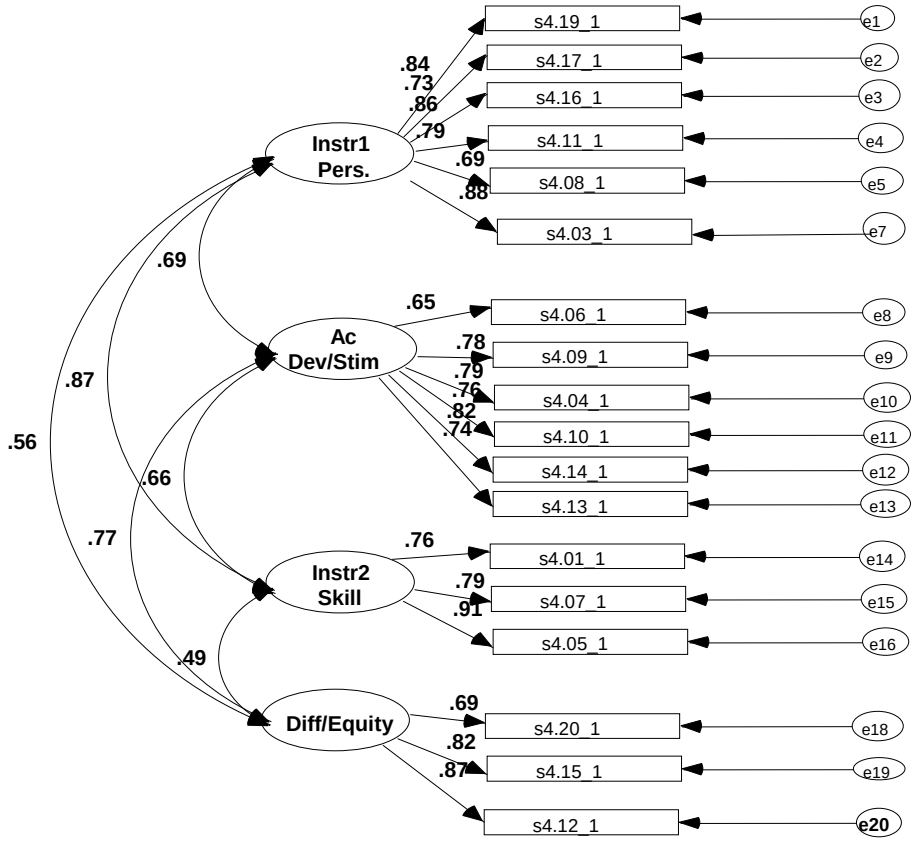
Satisfaction **n = 341** **Total Variance Explained = 68.1%**	**Factor 1 Instructor's Personable Aspects**	**Factor 2** **Academic Development**	**Factor 3 Course Difficulty / Equity**	**Factor 4** **Instructor's Skills/Ability**
Concern shown by instructor	.875			
Opportunities to discuss progress	.859			
Friendliness of instructor	.813			
Fairness of instructor*	.753			
Availability of instructor outside class	.666			
Fairness of marking/grading system*	.623			
Helpfulness of instructor	.543			
Attitude of instructor				
Personal attention provided				
Way instructor responds when you ask for help				
Degree to which you found course interesting		.888		
Found intellectually stimulating		.844		
Relevance to future occupation		.623		
Contribution to academic development		.616		
Nature of material covered**		.560		
Degree of challenge †		.558		
Required workload**			.881	
Level of difficulty*			.828	
Amount of study relative to grades*			.678	
Instructor's knowledge of subject matter				.814
Sequence of topics presented				.791
Ability of instructor to explain things clearly				.575
Courteousness of instructor				
*Enthusiasm of instructor****				
Eigenvalue	**8.208**	**2.376**	**1.409**	**.994**
Percent of variance explained	**43.2**	**12.5**	**7.4**	**5.0**
Cronbach's alpha	**.9102**	**.8668**	**.7983**	**.8120**

* Denotes the item has been reinstated
** Denotes the item has been re-worded
*** Denotes item was re-introduced for CFA
† Denotes the item is loading under a different factor
Items in italics indicate that they were discarded or not included for the second EFA



Number of Responses	874		
Chi-Square Value	163.243	**RMSEA**	0.057
Degrees of Freedom	43	**TLI**	0.993
p-value	0.000	**NFI**	0.995
Cmin/df	3.796	**CFI**	0.996
GFI	0.959	**Pclose**	0.113
AGFI	0.936	**AIC**	257.243

Figure 1: Structural Equation Model for Perceived Learning



Number of Responses	874		
Chi-Square Value	316.83	**RMSEA**	0.046
Degrees of Freedom	112	**TLI**	0.974
p-value	0.000	**NFI**	0.971
Cmin/df	2.829	**CFI**	0.981
GFI	0.962	**Pclose**	0.875
AGFI	0.942	**AIC**	434.83

Figure 2: Structural Equation Model and Fit Indices for Student Satisfaction

Table 3: Pearson Moment Correlations for the Relationship between Various Components of Student Satisfaction (N=723)

Variable		Sat. Univ General	Sat. Life General	Sat. InstP	Sat. AcDev	Sat. Diff/Eq	Sat. InstS
Satisf. University in General	Pearson Correlation	1	.558**	.212**	.201**	.208**	.137**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
Satisf. Life in General	Pearson Correlation	.558**	1	.131**	.159**	.185**	.090*
	Sig. (2-tailed)	.000	.	.000	.000	.000	.016
Satisf. Instructor - Personable	Pearson Correlation	.212**	.131**	1	.568**	.478**	.781**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
Satisf. Academic Dev.	Pearson Correlation	.201**	.159**	.568**	1	.636**	.539**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
Satisf. Difficulty/Equity	Pearson Correlation	.208**	.185**	.478**	.636**	1	.395**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
Satisf. Instructor-skills	Pearson Correlation	.137**	.090*	.781**	.539**	.395**	1
	Sig. (2-tailed)	.000	.016	.000	.000	.000	.

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).