

**Measuring Student Experience: Relationships between  
Teaching Quality Instruments (TQI) and Course Experience  
Questionnaire (CEQ)**

**Joe Hirschberg, Jenny Lye, Martin Davies &  
Carol Johnston**

Dec 2011

Research Paper Number 1134

ISSN: 0819-2642

ISBN: 978-1-921856-73-0



**Measuring Student Experience:  
Relationships between Teaching  
Quality Instruments (TQI) and  
Course Experience Questionnaire  
(CEQ)**

**Final Report**

**2011**

**Lead Institution:**

**The University of Melbourne**

**Partner Institutions:**

**Australian Council for Educational  
Research  
Flinders University  
Graduate Careers, Australia  
University of Tasmania  
University of Wollongong**

**Report Authors:**

**Joe Hirschberg  
Jenny Lye  
Martin Davies  
Carol Johnston**

Support for this project has been provided by the Australian Learning and Teaching Council Ltd, an initiative of the Australian Government. The views expressed in this report do not necessarily reflect the views of the Australian Learning and Teaching Council or the Australian Government

This work is published under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Australia Licence. Under this Licence you are free to copy distribute, display and perform the work and to make derivative works.

**Attribution:** You must attribute the work to the original author and include the following statement: Support for the original work was provided by the Australian Learning and Teaching Council Ltd, an initiative of the Australian Government.

**Noncommercial:** You may not use this work for commercial purposes.

**Share Alike:** If you alter, transform, or build on this work, you may distribute the resulting work only under a licence identical to this one.

For any reuse or distribution, you must make clear to others the licence terms of this work.

Any of these conditions can be waived if you get permission from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/3.0/au/> or send a letter to Creative Commons, 453 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Requests and inquiries concerning these rights should be addressed to the Australian Learning and Teaching Council, PO Box 2375, Strawberry Hills NSW 2012 or through the website: <http://www.altc.edu.au>

2011

ISBN 978-1-921856-73-0



# Contents

<b>List of Tables</b> .....	<b>iv</b>
<b>Table of Figures</b> .....	<b>v</b>
<b>Acknowledgements</b> .....	<b>viii</b>
<b>Executive Summary</b> .....	<b>1</b>
<b>Chapter 1. Introduction</b> .....	<b>3</b>
<b>Chapter 2. A review of influences on student evaluations of teaching in tertiary institutions</b> .....	<b>6</b>
2.1 Introduction .....	6
2.2 Purpose of collecting students rating data .....	6
2.3 Sources of contention in relation to the use of ratings data .....	7
2.4 Influences on student rating data .....	8
2.5 Instructor-level determinants .....	9
2.6 Student-level determinants .....	10
2.7 Subject-level determinants .....	12
2.8 Discussion .....	13
<b>Chapter 3. A systematic analysis of quality of teaching indicator surveys found in Australian tertiary institutions</b> .....	<b>14</b>
3.1. Introduction .....	14
3.2 The Data .....	15
3.2.1 The Institutional Practices .....	15
3.2.2 The Question Types .....	15
3.3 A Descriptive Analysis of Surveys Used by Institutions .....	18
3.3.1 The Classification by Focus on Student Learning .....	19
3.3.2 A Cluster Analysis of the Universities.....	20
3.4. An Analysis of Questions by Type. ....	22
3.4.1 Question Type Usage by Institutional Grouping .....	22
3.4.2 A Cluster Analysis of the Question Types .....	23
3.5. Conclusions.....	24



<b>Chapter 4. The matching of quality of teaching indicators and the course experience questionnaire.....</b>	<b>26</b>
4.1. Introduction .....	26
4.2. Description of Universities.....	27
4.3. Description of Surveys .....	28
4.3.1 Australian TQI Surveys as described in Chapter 3.....	28
4.3.2 CEQ.....	28
4.4 Results of analysis of New TQI and the CEQ .....	29
4.4.1 Item Response Modelling.....	30
4.4.2 Factor Analysis.....	31
4.4.3 Cluster Analysis.....	33
4.4.4 Regression Analysis .....	34
4.5. The analysis of the institution specific questionnaires.....	35
4.5.1 Introduction.....	35
4.5.2 Investigating construct invariance in institutional specific TQIs.....	36
4.5.3 Flinders University .....	38
4.5.4 The University of Melbourne.....	40
4.5.5 University of Tasmania .....	42
4.5.6 University of Wollongong.....	43
4.6 Discussion .....	45
<b>Chapter 5. The analysis of the response to the Course Experience Questionnaire (CEQ) as determined by the history of subjects taken by the respondents.....</b>	<b>47</b>
5.1 Introduction .....	47
5.2 Data.....	48
5.2.1 CEQ/GDS data .....	48
5.2.2 The Quality of Teaching (QOT) survey data.....	51
5.2.3 Who responds to the CEQ? .....	54
5.3. Analysis of the response to the Good Teaching Scale (GTS) of the CEQ ...	57
5.3.1 The Model Specification .....	57



5.3.2 Different Fixed Effects used for Control.....	58
5.3.3 Interpretation of the Estimated Coefficients.....	59
5.4 The student experience influence on the response to other CEQ scales. ....	64
5.5 Discussion .....	65
<b>Chapter 6. Discussion .....</b>	<b>68</b>
<b>References.....</b>	<b>71</b>

## List of Tables

<b>Table 3.1:</b> A Comparison of SLQ by Institution .....	19
<b>Table 3.2:</b> Typical Surveys by Cluster Sets membership in the 5 and 12 cluster cases.....	22
<b>Table 3.3:</b> The percentage of each group that uses each QT, ranked by frequency of all institutions use of the questions. ....	23
<b>Table 4.1:</b> New TQI administered at four universities.....	28
<b>Table 4.2:</b> The Good Teaching Scale .....	28
<b>Table 4.3:</b> The Generic Skills Scale .....	29
<b>Table 4.4:</b> The Overall Satisfaction Item .....	29
<b>Table 4.5:</b> Fit statistics for the new TQIs versus CEQ.....	32
<b>Table 4.6:</b> The regression parameter estimates of the responses to the new TQI on the CEQ questions. ....	35
<b>Table 4.7:</b> Institution-specific items .....	37
<b>Table 5.1:</b> The distribution of the number of courses per reporting student. ....	48
<b>Table 5.2:</b> The distribution of courses by faculty and year taken by reporting student. ....	49
<b>Table 5.3:</b> The classification of CEQ questions by scale.....	50
<b>Table 5.4:</b> The questions used on The University of Melbourne QOT.....	51
<b>Table 5.5:</b> The results of the tests of hypotheses concerning the parameters estimated in model (1). ....	52
<b>Table 5.6:</b> A weighted regression over the average response to question #2 by department.....	54
<b>Table 5.7:</b> The results of a probit analysis to determine the factors that influence the proportion of subject's enrolment that ultimately respond to the CEQ. (prob $\chi^2$ )	



indicates the significance of the test that the parameter estimate is equal to zero)	55
<b>Table 5.8:</b> The model definitions by the fixed effects accounted for (see appendix F in the Resource Document for this Report for details)	58
<b>Table 5.9:</b> The parameter estimates and the F-test parameters for the Mixed models estimated for the Good Teaching Scale (GTS) using only Australian Permanent resident students.	60
<b>Table 5.10:</b> $R^2$ for each model/dependent variable combination.	64
<b>Table 5.11:</b> Coefficient estimates across the dependent variables using model #1 which uses the course code from the University of Melbourne for the definition of the fixed effects.	65

## Table of Figures

<b>Figure 3.1:</b> The Dendrogram of the hierarchical cluster analysis of Australian universities created using a complete linkage method. The C–5 and C–12 columns indicate the membership in either the 5 group clustering or the 12 group clustering.....	21
<b>Figure 3.2:</b> The Dendrogram of the Hierarchical Cluster Analysis of the QTs based on a complete (or furthest neighbour) linkage method.....	24
<b>Figure 4.1:</b> The average score for the new TQI and CEQ questions. ....	29
<b>Figure 4.2:</b> Item Person Map for new TQI and CEQ for four Universities .....	30
<b>Figure 4.3:</b> Fit (mean square): CEQ and New TQI items .....	31
<b>Figure 4.4:</b> Scree Plot of Eigenvalues by number .....	32
<b>Figure 4.5:</b> The Component Plot of the first two Factors (Using a varimax rotation) .....	33
<b>Figure 4.6:</b> Dendrogram using Complete Linkage and the Chi-square test for the equality of two sets of frequencies as the measure of distance.5. ....	34
<b>Figure 4.7:</b> Item person map: CEQ and Flinders University items .....	38
<b>Figure 4.8:</b> Fit (mean square): CEQ and Flinders University items .....	39
<b>Figure 4.9:</b> Scale mean score relationships: CEQ and Flinders University .....	39
<b>Figure 4.10:</b> Item person map: CEQ and The University of Melbourne items .....	40
<b>Figure 4.11:</b> Fit (mean square): CEQ and The University of Melbourne items .....	41
<b>Figure 4.12:</b> Scale mean score relationships: CEQ and The University of Melbourne .....	41
<b>Figure 4.13:</b> Variable map: CEQ and University of Tasmania items .....	42



<b>Figure 4.14:</b> Fit (mean square): CEQ and University of Tasmania items.....	43
<b>Figure 4.15</b> Scale mean score relationships: CEQ and University of Tasmania....	43
<b>Figure 4.16:</b> Variable map: CEQ and University of Wollongong items.....	44
<b>Figure 4.17:</b> Fit (mean square): CEQ and University of Wollongong items.....	45
<b>Figure 4.18:</b> Scale mean score relationships: CEQ and University of Wollongong	45
<b>Figure 5.1:</b> The interrelationship of the data used to perform the analysis. ....	49
<b>Figure 5.2:</b> The Dendrogram of the responses to the CEQ as reported by students at The University of Melbourne .....	50
<b>Figure 5.3:</b> A set of scatter plots of the department averages.....	53
<b>Figure 5.4:</b> The histograms of the number of categories by size of the seven fixed effects (note that Duties and Industry exclude the unknown category).....	59
<b>Figure 5.5:</b> The distribution of all marks issued to the students in this sample. ....	61



## Project Team

### Australian Council for Educational Research

Dr Hamish Coates (Senior Research Fellow, Australian Council for Educational Research)

### Flinders University

Associate Professor Heather Smigiel (Director of Academic Development)  
Staff Development and Training Unit

### Graduate Careers Australia

Ms Cindy Tilbrook (Executive Director, Graduate Careers Australia)

Mr. Bruce Guthrie (GCA Research Manager, Graduate Careers Australia)

### University of Tasmania

Mr. Nigel Ewan, Deputy Academic Registrar

### University of Wollongong

Professor Sandra Wills, Director, Centre for Educational Development & Interactive Resources



## Acknowledgements

This report was prepared under the grant CG7-384 from the Australian Learning and Teaching Council.

In particular, we wish to thank Mr. William Jones of the Planning Department of the University of Melbourne for help in preparing the data for this analysis of the CEQ responses from The University of Melbourne. In addition to the named participants in this study, we also wish to thank attendees at the following conferences where parts of this report were presented. The *World Universities Forum 2011*, Hong Kong, January 14<sup>th</sup>, 2011, The 8<sup>th</sup> Annual Hawaii International Conference on Education, Honolulu, HI, January 10, 2010; The Forum on Quantitative Analysis Of Teaching And Learning In Higher Education In Business, Economics And Commerce, The University of Melbourne, February 12, 2010; The 15th Australasian Teaching Economics Conference, Hamilton New Zealand 29<sup>th</sup> of June 2010; The Forum on Quantitative Analysis Of Teaching And Learning In Higher Education In Business, Economics And Commerce, The University of Melbourne, February 6, 2009; an invited seminar, Department of Economics, Southern Methodist University, Dallas, TX, January 8<sup>th</sup>, 2009; and the 7<sup>th</sup> Hawaii International Conference on Statistics, Mathematics, & Related Fields 2009, January 15, 2009.

We also wish to thank those lecturers and students in the subjects at Flinders University, the University of Tasmania, the University of Wollongong and The University of Melbourne that were chosen to survey for the results presented in Chapter 4. In addition we wish to thank those institutions that have provided the teaching quality instruments used in the analysis that we discuss in Chapter 3.

The cover photograph was taken of participants at the Forum on The Quantitative Analysis of Teaching and Learning in Higher Education in Business, Economics and Commerce, held on February 11, 2011 at The University of Melbourne.



## Executive Summary

The results of the Course Experience Questionnaire (CEQ) in Australian tertiary institutions have been available for a number of years and have provided the administration of these institutions with valuable information as to students' perceptions of their courses. In addition to the CEQ, all of these institutions survey their students at the subject level. The purpose of this study has been to determine the degree to which the responses recorded on these subject level Teaching Quality Indicators (TQI) are related to the CEQ.

In the analysis, we have used two approaches to cast some light on the degree to which the responses on the TQI can be used to anticipate and establish responses on the subsequent CEQ.

Prior to performing these studies, we first sampled the extensive literature that describes influences on responses to TQIs. From this survey, we found that there are many factors which influence students' responses, other than a specific lecturer's performance in class. We have used the literature on the TQIs because there is almost no literature concerning students' responses to course level surveys such as the CEQ. Nonetheless we have used the results of this survey to determine how we can control for these factors when modeling responses to the CEQ.

There were two major hurdles which needed to be addressed in the performance of this analysis across institutions.

- First, there is no consistent TQI for all institutions. Most of these surveys have been assembled using questions with different origins that have varied over time and that are individual to each institution.
- Secondly, even if the TQIs were constant and consistent across all institutions, it would be necessary to account for any other factors, such as the year level and the field of study, that may influence the specific subject TQIs and the overall course evaluation.

Our first analysis addresses, in a systematic manner, the first difficulty by investigating the nature of the TQIs in use in Australian tertiary institutions. Through the use of a website designed as a clearing house for information on these surveys, we have collected examples of the TQIs from 39 higher education institutions in Australia. We then categorised these different instruments to develop a composite new TQI survey that captures the characteristics of those in use.

Using the new TQI, we then compared the responses of students at the four participating institutions (Flinders, Melbourne, Tasmania, and Wollongong). The students were asked to respond to a survey that included questions relating to teaching, generic skills and overall course satisfaction (from the CEQ), questions from the local TQI, and questions from the new TQI based on the composite question types found in our analysis of all Australian TQIs. We found that few of the institution-specific TQI's elicited responses that matched the CEQ and that even the new TQI did not provide many solid matches. Thus we concluded that, in most cases, the items on the TQI and the CEQ are measuring different factors.

For the CEQ responses from The University of Melbourne, we matched the students' experiences in order to determine the degree to which their responses could have been anticipated. From this analysis we found that those students who completed the CEQ were most responsive to situations where their expectations are



not met. We also found that characteristics of subjects that influence TQI responses are also factors that influence CEQ responses. We found that older students, those that receive higher marks than their peers, and those who study full-time in their last year, are more likely to rate a course more highly. Whilst students who take more than one course, who are in subjects with higher than average enrolments, and that received lower than average TQI ratings, are less likely to rate their course highly.

The conclusion from this study is that TQIs at different institutions are not designed in a consistent manner and that only a small portion of the CEQ responses could be predicted by these TQI. However, just as with the studies of the TQIs, we can establish that course characteristics such as: the level of the degree, the Faculty and Department in which the course was taken, the course description, the industry and duties of those who have found employment after completing their course, all strongly influence the CEQ.



## Chapter 1. Introduction

Previously quality assurance in Australian higher education relied on institutional self-assessments and discipline reviews, however, recent trends have placed a much greater emphasis on quantitative measures of institutional performance. One of the most important developments has been the incorporation of data from the Course Experience Questionnaire (CEQ) (Ramsden, 1991a, 1991b; Ramsden & Entwistle, 1981) into national benchmarking and funding decisions. As the demands for institutional accountability continue to escalate, it is imperative that we work to better understand the data on which we base our decisions and how such data relates to educational change processes.

Since 1993, the CEQ has been mailed to all graduated students in Australia who have completed a qualification in the previous year. It has been adapted for use with distance education students (Lawless & Richardson, 2002; Richardson, 2003) and has also been evaluated in specific disciplines, notably Business (Athiyaman, 2001); Medicine (Broomfield & Bligh, 1998; Lyon & Hendry, 2002); Nursing (Trigwell & Prosser, 1991); Occupational Therapy (Sadlo & Richardson, 2003); Leisure, Sport and Tourism (Downie & Moller, 2005); and Accounting (Byrne & Flood, 2003). A number of studies have shown that the instrument and its scales have reasonably robust psychometric properties (Eley, 2001; Richardson, 1994; Trigwell & Prosser, 1991). Moreover, the CEQ is now also used as a key element in determining Commonwealth Learning and Teaching Performance funding.

Despite the role it has assumed in the evaluation of higher education quality, the CEQ has limitations. The lagging and aggregate nature of the data make it difficult for institutions to use CEQ data alone in their internal continuous, locally-responsive quality improvement activities. The key aim of the CEQ is to measure student perceptions of their courses of study and to assess differences between academic organizational units in terms of those perceptions (Lizzio, Wilson, & Simons, 2002; Ramsden, 1991b). Whilst often of primary interest to institutions and teaching staff, the CEQ is not designed to measure student perceptions of individual lecturers or units of study.

In order to gain an understanding of student perceptions of individual lecturers or units of study nearly all Australian higher education institutions have, in recent years, developed institution-specific instruments and surveys to provide context-relevant data. These Teaching Quality Instruments (TQIs), with names such as QOTs, SETs, LETs, TEVALs,<sup>1</sup> have grown to play an important role in quality assurance in Australian higher education. TQIs are used in almost all of the 43 public and private tertiary institutions in Australia, and in 16 their use is compulsory. In the remaining institutions, their use is virtually mandatory for promotion and advancement purposes.

Whilst TQI surveys are subject to influences outside the control of lecturers (Davies, Hirschberg, Lye, McDonald, & Johnston, 2005a, 2005b, 2006), they are more sensitive to specific aspects of local educational contexts than the CEQ. Like the CEQ, virtually all tertiary institutions rely on the TQI for regular annual performance measures in relation to student satisfaction. However, they rely on them for different

---

<sup>1</sup> A complete list of names for these surveys is listed in Table A.1 in the Appendix provided in the Resource Document for this project.



reasons. Where the CEQ allows for cross-institutional measurement of the quality of courses; TQIs allow for intra-departmental and university-level comparisons of teaching quality.

A closer alignment of the CEQ and TQIs is an important step in the direction of more valid and sensitive data in the quality of university education in Australia. It will support a movement towards developing a national approach to quality of teaching measurement that is robust enough for cross institutional comparisons whilst at the same time providing individual institutions with information that is useful in improving teaching practice.

Although the CEQ and various TQI measures play important and complementary roles in higher education quality assurance, little is known about the relationships between these instruments. It is reasonable to assume that the end-of-course CEQ and the subject-level TQI instruments are related. Where the CEQ evaluates students' perceptions of their courses, TQIs are designed to measure students' perceptions of teaching at the subject level. It is unlikely (although not inconceivable) that students would evaluate teaching and units of study using a TQI in an entirely different way from how they evaluate courses using the CEQ. Whilst it is reasonable to assume that CEQ and TQI measures are related, this assumption has not yet been empirically tested. Without such evidence, it is difficult to determine how institutions should react to this data in order to improve learning and teaching practice.

This report outlines the results of the research to determine the extent to which the Course Experience Questionnaire can be anticipated by subject specific questionnaires like TQIs. As such, the research required the consideration of a number of questions.

- Is there a standard subject questionnaire used in Australian tertiary institutions and how do students respond to a standardized questionnaire as compared with the CEQ?
- To what extent do previous subject experiences, including the subject specific average responses on teaching questionnaires, other subject characteristics, student performance, and student characteristics influence responses to the CEQ?

The primary conclusion from the research carried out in this study is that there is little correspondence between the typical subject level survey and the CEQ even when these questions are asked of students at the same point in time, as done for the surveys described in Chapter 4.

The other result is that the cumulative impact of other experiences on the student whilst at the institution may have more influence on responses to a CEQ questionnaire than results measured by subject specific questionnaires. In the case of the subject questionnaires, the degree to which responses are *below* the expected level for the subject are far more important than any improvement in these values. This outcome, along with an analysis of the characteristics of those students who do fill out the CEQ, may be most influential for the institution.

In this research, we needed to perform a number of steps to make the connections between these surveys. In this report we develop two avenues of research. First is the determination of the nature of subject specific questionnaires. Second is the



examination of the impact of aggregate responses to these subject experience questionnaires on those same students' responses to the CEQ.

Our research, therefore, investigates the relationship between the subject level experiences of students and their subsequent responses on the CEQ at a major Australian tertiary institution. This was done by constructing a history of the subjects taken by those students who completed the CEQ. A series of descriptive measures for each of these students' subjects were also developed, such as: enrolment numbers, the distribution of marks and the average responses on the subject TQIs. We then used these descriptive measures, along with the students' personal experiences and their responses to the Graduate Destination Survey (GDS) to determine how these factors may have influenced their evaluation of their course of study through the CEQ.

We have found that the age of the student, their relative mark, their class rank in the last year, and the total number of subjects they have taken at the university all have a positive impact on their evaluation of the teaching quality, whilst a positive change in their average marks in the last year and the number of completed courses all have a negative impact. On examining the characteristics of the subjects they took we have found that larger class sizes, a large proportion of close marks given in the subject (not necessarily to them), and lower quality of the subjects as measured by the TQIs for the subject when compared to the average for the subject, all resulted in a lower value for the responses. In addition, we also find evidence that the average TQI scores for the subjects taken have a positive impact on the teaching scores.

The Australian government encourages all nationally funded tertiary institutions to survey their graduates using a common instrument. The Course Experience Questionnaire (CEQ) is designed to gather information on a number of aspects of student's experiences during their course of study at the institution. Up until 2009, parts of the results of these surveys were used to determine the funding of special grants for institutions that score well. In addition to these end-of-course surveys, most institutions in Australia conduct end-of-semester surveys or TQIs within each subject taught. In this analysis we survey students in a number of institutions to establish the relationship between the responses on the locally administered surveys (TQIs) and the responses on the CEQ survey in order to establish the degree to which the indications of good teaching as defined in the CEQ coincide with the positive responses elicited from the TQI. Whilst all tertiary institutions in Australia use the same Course Experience Questionnaire (CEQ), for the internal evaluation of teaching they use subject specific TQIs. This research has also included an classification of the TQI questions used in Australian universities. This classification has enabled us to explore how internal universities' surveys are similar to or different from each other. We have also sought to establish whether there are particular patterns questioning. This analysis can then be used by universities to determine how their surveys compare with their peer institutions and other institutions across Australia.



## Chapter 2. A review of influences on student evaluations of teaching in tertiary institutions.

This review of the literature on influences to student evaluations is intended to provide the background to how student surveys such as the CEQ and TQIs are influenced by a series of factors which may not correspond to the aim of measuring the quality of instruction. In this chapter we survey the literature where factors beyond the control of the instructor have been found to contribute to the subject specific Teaching Quality Instrument results.

### 2.1 Introduction

In Australia, as in Britain, there is an impetus to quantify the quality of teaching in order to compare institutions and departments and to compare individuals at the personal level. The use of surveys that collect information on student perceptions of teaching is now widespread. Whilst the subject of this report is not to debate the best way to proceed in order to improve tertiary education, it is appropriate to explore whether student perception survey data are adequately robust for this purpose.

The widespread use of surveys of student perceptions of their teaching and learning experience has prompted an almost equally widespread investigation of what these surveys are actually measuring. There are literally thousands of references to research on student ratings of teaching, most in the last two decades. Most research has been conducted in the United States (see for review Cashin, 1995, and Marsh 1987. There are a handful of studies emanating from Australia (Bedggood & Pollard, 1999; Haynes, 2002; Marsh, 1987; Marsh & Bailey, 1993; Neumann, 2000; Wagner, 1999; Worthington, 2002) and Europe (Husbands, 1996, 1997; Husbands & Fosh, 1993; Shevlin, Banyard, Davies, & Griffiths, 2000). Few studies have been conducted over an extended period of time (Haynes, 2002; Marsh & Bailey, 1993; Marsh & Hocevar, 1991; Ting, 2000). The issue of the validity of student evaluations of teaching is a matter of some controversy with evidence both supporting their continued use and evidence recommending their discontinuation. There is no sign of the controversy abating (Abrami, d'Apollonia, & Cohen, 1990; Boice, n.d.; Cashin & Downey, 1992; Dwinell & Higbee, 1993; Greenwald, 1997; Hepworth & Oviatt, 1985; McKeachie, 1997; McKeachie & Lin, 1979; Smith, 2004; Solas, 1990).

The first part of this chapter provides a broad overview of the existing literature in relation to student surveys and identifies potential biases and systematic influences. The second part of this chapter reports an analysis of Australian student ratings in a department of economics collected over a ten-year period. Outcomes of this analysis and the implications for teaching practice are then discussed.

### 2.2 Purpose of collecting students rating data

Student rating surveys in tertiary institutions have been used to gain diagnostic feedback on teaching effectiveness (Marsh & Bailey, 1993; Marsh & Roche, 1993) in assessing the particular instructor characteristics that might assist in improving individual teaching performance. They have been used for administrative decision-making (McKeachie, 1997; Simpson & Siguaw, 2000; Wolfer & Johnson, 2003) for summative and formative purposes, as a basis for hiring, confirmation and



promotion or in matching appropriate teachers to appropriate courses (Williams & Ceci, 1997; Wilson, 1998). They have also been administered to fulfill demands for the accountability of institutions (Williams & Ceci, 1997) so that cross institutional comparisons can be made (Access Economics 2005). They have also been available in some institutions for students to use in informing their subject selection decisions (Marsh 1987). The appropriateness of using this rating data in these ways is a matter of deep controversy.

### 2.3 Sources of contention in relation to the use of ratings data

Several studies have confirmed the validity and reliability of student ratings in the USA (Bosshardt & Watts, 2001; Wachtel, 1998), in Australia (Marsh, 1987), in Hong Kong (Ting, 2000) and in Europe (Byrne & Flood, 2003; Husbands, 1996, 1997; Husbands & Fosh, 1993). Despite this there is a great deal of debate about the usefulness of these ratings. Some argue that the precise numerical scores generated in student evaluations 'imply a level of measurement that simply does not exist' (Wolfer & Johnson, 2003), p.117) and that it would be better if numerical scores were replaced with categories such as "exceptional", "adequate" and "unacceptable" (d'Apollonia & Abrami, 1997), more global measures of assessment (Cashin & Downey, 1992) or a range of scores (Neumann, 2000). Others suggest that the fine measurements generated do not allow discrimination between good and bad teaching and that "cutoff" points would be a more sensible approach (McKeachie, 1997). Still others claim that the use of student evaluations for administrative purposes is a misappropriation of the data (Bedggood & Pollard, 1999; Sheehan, 1975) in part as students themselves may have little idea that such use is made of their evaluations (Dwinell & Higbee, 1993).

A common form of instrument has been the student questionnaire that uses Likert-type scales as a measure of student opinion. Often, the ratings from such instruments are accepted without question, even though it is not known whether all students interpreted the items on such questionnaires in a similar way, or whether students were consistent in their interpretation of the rating scale. Students may reinterpret the meaning of the items in ways which may be quite different from the intentions of the questionnaire designer or researcher (Low, 1999). Block (1998) complemented the administration of a student questionnaire with a small number of student interviews and found a high degree of variance in the responses of students to questionnaire items and to the ratings they had given to those items.

The method of survey administration is also an area of debate and one that can affect the ratings given by students. Dwinell and Higbee (1993) found that a majority of students believed their anonymity was assured when they provided ratings of their instructors, whilst a Canadian study (Fries & McNinch, 2003) revealed that the anonymity had some effect on students' ratings of teachers. Students who had signed their ratings forms gave higher ratings to their teachers.

Kolitch and Dean (1999) point to assumptions about curriculum, instruction, evaluation and student-teacher relationships underlie the items of a typical rating instrument. Even though a rating instrument might claim to be representative of all conceptions of teaching, such an instrument is more consistent with a transmission paradigm of teaching. Solas (1990) recognised that data obtained through surveys are confined to a small number of dimensions, often "more meaningful to the surveyor than the individual being surveyed" (p.152). Instead Solas used a repertory grid which enabled students to express their own opinions, in their own



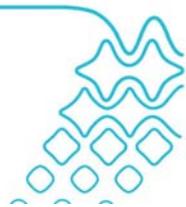
terms, thereby providing more detailed information than possible through a questionnaire (Solas, 1990). Other sources of data, such as student journals (Wagner, 1999) and teacher interviews (Ballantyne, Bain, & Packer, 1999; Hativa & Raviv, 1993; McCormick, 1996) have also been employed to gather student perceptions of their teachers.

Furthermore, the usefulness of student ratings in improving teaching is questioned. Students do not have the knowledge necessary for appropriate rating of teaching (Simpson & Siguaw, 2000), nor are they necessarily best judges of their instructors' performance (Casey, Gentile, & Bigger, 1997), viewing it 'from very limited or even tainted perspectives' (p. 472). Students also consider the provision of feedback (such as evaluation surveys) as a chore (Simpson & Siguaw, 2000). For their part, though, students perceive that their evaluations of instructors are heeded and that instructors do change their behaviour as a consequence (Dwinell & Higbee, 1993) even though later cohorts of students have no means of comparison, so are unlikely to be aware of such changes (Haynes, 2002). Indeed, information gathered from such evaluations is of little use to rectify problems for current classes and may be inappropriate for subsequent classes if the evaluations are conducted at the end of a term or semester (McKeachie, 1997). The usefulness of student evaluations can be assured only if data gathering, reporting and interpretation are carried out in a careful, considered manner (Casey et al., 1997) For instance, the influence of different teaching contexts should be considered (Neumann, 2000) and scores should be weighted according to discipline area, class size and other factors (Wolfer & Johnson, 2003).

## 2.4 Influences on student rating data

The majority of studies in this area have gathered quantitative data from rating surveys that use Likert-type rating scales. There are a number of factors that are explored in terms of the potential effect of these scales on survey responses. These can be broadly grouped in terms of the influence of: teacher or instructor-level determinants, student-level determinants and subject-level determinants. Variables in the teacher-level category include the instructors' use of class time, their availability outside class time, how well they assess student learning or understanding, their concern for students' welfare and performance, the extent to which they emphasise analytical or critical skills, their preparedness, their tolerance of alternative viewpoints in class. Student-level determinants include the reasons for taking the course, the class-level of the respondent, the effort students expend in the subject, age of the student, ethnicity and student gender. Subject-level determinants include when the subject is offered, whether it is required for the degree or a prerequisite subject, the level, the perceived difficulty, grade inflation and the size of the class.

Questions that are raised address in the literature include how these factors influence student ratings of their teachers, and which factors are the most crucial; whether the measures are biased or whether they represent an accurate indication of instructor teaching effectiveness; what the relationship is between student measures of teachers' performance and teachers' measures of student performance, for example whether good instructor ratings can be "bought" by giving students good grades. Studies are reviewed in the following section in terms of instructor-level determinants, student-level determinants and course-level determinants.



## 2.5 Instructor-level determinants

Several key studies since the 1970s have outlined the instructor-level determinants that are key to being a “good” teacher. Characteristics that students regard as meeting the criteria for an “ideal” teacher map directly to attributes that closely match traditional criteria of teaching competency (Pozo-Munoz, Reboloso-Pacheco, & Fernandez-Ramirez, 2000). Pozo-Munoz et al. (2000) identify the four main factors contributing to perceptions of good teaching as teaching competency, teaching qualities, teacher appearance, and directiveness. Analysis of the data shows that the most crucial attributes are related to teaching competency (which included “expertise”, “informed”, “clear”, “able to motivate” among 16 other attributes) and the least important related to directiveness (which included “prestigious”, “attractive”, “kind” (Pozo-Munoz et al., 2000) and other characteristics relating to personality and personal appearance. Further factors include clarity in teaching (Hativa, 1998; Ting, 2000); showing good management of student behaviour; demonstrating excitement and interpersonal skills (Lowman & Mathie, 1993); being able to provide intellectual stimulation (Ting, 2000); showing a caring nature and being systematic (Brown & Atkins, 1993); showing respect for students and being organised and having good presentations skills (Patrick & Smart, 1998).

Aigner and Thum (1986) found that 65% of the variation in the evaluation ratings could be explained by instructor-specific characteristics. The enthusiasm of the teacher, the level of interest stimulated, and the teachers’ interaction with students were among the most important instructor attributes.

DeCanio (1986) found that communication skills and the level of organisation of the lecturer to be the most important characteristics. Boex (2000) used factor analysis to define six composite attributes of instructors: presentation skills, organisational skills and clarity of expression, how the instructor used grading and assignments, intellectual or scholarly capabilities, the ability to interact well with students, and the ability to motivate students. He concludes that student evaluations were most likely to be influenced by the lecturer’s organization, clarity, ability to motivate students, and the grading practices used by the lecturer, whilst Mehdizadeh (1990) observed that expected grades, usefulness of supporting materials and the lecturer’s help outside class to be critical to the evaluation results. In a study of the teaching of economics, students evaluated their instructors’ effectiveness on the bases of presentation, organisation, knowledge, accessibility, responsiveness and enthusiasm (Gokcekus, 2000).

Williams and Ceci’s (1997) summary of previous findings indicated that, whilst ratings are reliable and do not change with student age, they are significantly influenced by instructors’ personal characteristics such as variations in voice patterns, warmth, supportiveness, dominance and confidence rather than attractiveness, gender or age. Whereas Shevlin et al.’s (2000) study demonstrates that lecturer “charisma” is a significant underlying variable influencing student evaluation rankings, and they have argued that—as a consequence—evaluation surveys are not accurate measurements of teacher effectiveness. Radmacher and Martin (2001) found that teacher’s extraversion was the only significant predictor of student evaluations after controlling for enrolment status, course grades and student ages. Felton, Mitchell and Stinson (2004) reach a similar conclusion in relation to professors rated as ‘sexy’. Students do appreciate instructors who are knowledgeable, warm, outgoing and enthusiastic (Murray, Rushton, & Paunonen, 1990). However, these same traits are likely to make the person a more effective



teacher, so that students are stimulated to greater achievement and learning. Other studies, however, show that student evaluations are not unduly influenced by the instructor's personality and popularity or ability to entertain (Costin, Greenough, & Menges, 1971; Marsh & Ware, 1982; McKeachie, 1978). If students feel they have learned they will give higher ratings. So the important factor is not how entertaining the instructor is: 'Neither the "stand-up comic" with no content expertise nor the "cold fish" expert with only content expertise receives the highest ratings consistently (Braskamp & Ory, 1994, p. 180; Murray et al., 1990)

Davies et al. (2007) found that, after accounting for a series of student and subject characteristics, there was a significant proportion of the variation in scores that were attributable to specific instructors and which would be due to the combination of teaching ability and instructor characteristics.

Overall evaluations tended to be negatively correlated with age and with years of teaching experience (Wolfer & Johnson, 2003). Studies of these variables indicated a negative correlation between ratings of teachers and their teaching experience in that assistant professors were rated higher than full professors. Also the title, degree and position of an instructor, i.e., teaching assistants, visiting professors, tenure-track assistant professors or tenured professors, do not influence students' evaluations of their instructors. Administrative experience, research publication and teachers' rank also had little effect on ratings (Ting, 2000).

The reputation of an instructor, if known by students before enrolling in a subject, was shown to influence student ratings: more highly reputed instructors received higher ratings than others (Griffin, 2000). This would indicate that the expectations of the student prior to taking the subject are one of the factors that influence their eventual evaluation. In yet another study, economics students gave higher ratings on all evaluated items of teaching effectiveness to those instructors who spoke English as a first language compared with those for whom English was not their native language (Bosshardt & Watts, 2001):

## 2.6 Student-level determinants

Given that student ratings may be made on the basis of features outside the lecturer's control, a number of studies have attempted to determine if student ratings do accurately measure teaching. The possibility of biases in student ratings has been subject to much study.

One key source of potential bias has been the influence of gender on teaching ratings, however results are inconclusive. McKeachie (1979) and Feldman (1993), for example, found no significant effect for gender on student evaluations and Cashin's 1995 review of the literature supported this view. Other studies, though, have reported some effect for gender. Basow and Silberg (1987) surveyed the perceptions of more than 1000 students from humanities, social sciences, natural sciences and engineering of the effectiveness of 16 pairs of male and female instructors who were matched on the basis of rank, type of course taught and years of experience. They found that both male and female teachers were rated as effective, but whereas female students rated male and female teachers similarly, male students gave significantly less positive ratings to female teachers. Basow (2000) also asked students to describe their "best" and "worst" teachers, and found that, overwhelmingly, female students chose a female teacher as "best". When describing "worst" teachers, however, no similar disproportion was evident. Wolfer



and Johnson (2003), in a study involving social work students, found that courses taught by female instructors were rated significantly higher than those taught by males. They suggested, however, that such a result might be explained by the fact that students tend to rate instructors of the same gender higher; approximately 90% of the participants in their study were female. In contrast, other studies have revealed that male instructors have been rated more highly than females. A review of research on college students' preconceptions of male and female college teachers revealed that in the majority of studies students' global evaluations of male and female teachers as professional were not different, though in a minority of studies male teachers received higher overall evaluations than did female teachers (Feldman, 1992). More recently, a study of 769 student evaluations of teaching in a Canadian university department of sociology and social studies found students rated male instructors more highly than female (Fries & McNinch, 2003).

Age of the student is another potential source of bias in student ratings. Centra (1993) found that the age of a student had no effect on student ratings. However, Worthington (2002) found that the age of the student did have some influence on the ratings given: students who were over the age of 30, and were also female, were more likely to assign a lower rating to the instructor.

The aptitude of students has been found to have some effect on the evaluations made of instructors. A study of economics subjects and teachers by Mason et al (1995) revealed that better students were tougher in rating the quality of the subject, but more lenient in their rating of lecturers.

Positive but low correlations have been reported between student ratings and expected grades. That is, students expecting high grades in a subject tended to give higher ratings than did students expecting lower grades (Aigner & Thum, 1986; Mehdizadeh, 1990; Millea & Grimes, 2002). Worthington (2002), investigating the influence of student characteristics on the probability of particular rankings of one lecturer in a finance subject, made similar findings. Higher ratings were likely to be given by students who were expecting a higher grade and were from a non-English speaking background.

Grimes, Millea & Woodruff (2004) examined the relationship between student evaluations of their instructors and the innate and personal psychological construct of locus-of-control, a construct which identifies an individual's belief in their control over the environment. Those, for example, who believe they have little control over personal outcomes are categorised as 'externally oriented', whereas 'internally oriented' are those who accept responsibility for such control. Grimes et al (2004) found that students identified as internally oriented are more likely to evaluate instructors highly than are externally oriented students who, believing that they had little personal control over their grade, blame outside factors for their performances and so assign lower ratings to the instructor.

Davies et al. (2007) conclude that, in the Australian context, the mix of country of origin and gender in particular subjects had a significant impact on TQI scores. In particular, in the economics subjects taught at the University of Melbourne, female students from the Indian sub-continent, China, Hong Kong, the UK and Singapore and male students from the UK and China were harder on the instructors than male Australian students with all other factors held constant.



## 2.7 Subject-level determinants

Most studies on the influence of class size on student perceptions of quality of teaching have found little, if any, effect on the evaluation of teaching effectiveness (Marsh, 1987). Although there is a tendency for smaller classes to receive higher ratings, it is a very weak inverse relationship (Cashin, 1995; Sixbury & Cashin, 1995). However, an Australian study conducted across four discipline groupings — humanities, sciences, social sciences and professionally oriented subjects - found that larger classes were rated lower (Neumann, 2000). Class size effect has also been found by (Feldman, 1984; Liaw & Goh, 2002).

Ratings of perceived teaching quality do not appear to be related to the time of day or part of the year when the course is conducted (Liaw & Goh, 2002), nor to when the ratings are conducted (Abrami, Leventhal, & Perry, 1982; Cuseo, 2002; Seldin, 1993; Feldman, 1979). Nevertheless, Cuseo (2002) suggests that rating surveys administered immediately after a final examination might have less validity overall, due to the preoccupation, anxiety or fatigue of the students at that time.

Ratings by individual students of teachers do not change over time, with student post-subject experiences, or with the increased maturity of students. In cases where students have given retrospective ratings of subjects completed up to five years earlier, substantial agreement has been found between the retrospective ratings and those made at the time of the subject completion (Feldman, 1989; Feldman, 1989; Overall & Marsh, 1980). Such findings refute the argument that, because they are immature, students will only come to appreciate subjects or instructors who were initially rated poorly with greater maturity.

Several studies have indicated the existence of a relationship between the nature of the discipline and student ratings. Feldman (1978) found that humanities and arts-type subjects receive higher ratings than social science-type courses and that these, in turn, received higher ratings than mathematics-type courses. Others have found similar results (Braskamp & Ory, 1994; Cashin, 1990; Centra, 1993; Marsh & Dunkin, 1992; Neumann, 2000; Sixbury & Cashin, 1995, 1995). Nevertheless, Cashin (1988), recommended caution in the interpretation of such findings, suggesting that, if it were the case that instructors in fields requiring quantitative reasoning skills were rated lower because students were less competent in those skills, this would introduce a bias in student ratings. Such a bias, Cashin argued, should be corrected. However, Marsh (1982), in a study designed to estimate the effects of normally confounding variables, concluded that the instructor is the primary determinant of the student rating, with an effect about five times as large as the effect of the subject.

Others (Saroyan & Snell, 1997) have suggested that the type of instructional method employed can influence students' evaluation of the instructor. In a Canadian medical school study of three lecturers described as 'content-driven', 'context-driven' and 'pedagogy-driven' (pp. 99-100), Saroyan and Snell (1997) found that the student-centred context-driven and pedagogy-driven lectures were more highly rated than the teacher-centred content-driven lecture.

The influence of grades on student evaluations is an important issue in the literature. Studies of a number of undergraduate courses found that the grading leniency of instructors, together with expected workload in a subject, influenced the ratings of students (Greenwald & Gillmore, 1997). It has been suggested that some lecturers



manipulate grades to obtain higher evaluation scores (Nelson and Lynch 1984; Zangenehzadeh, 1988, and Krautmann and Sander, 1999) Howard and Maxwell (1980, 1982) however found little causal evidence of grading leniency and student satisfaction. Conversely, in Hong Kong (Ting, 2000), a teacher who gave higher grades was likely to receive lower ratings from students. There, it is suggested that as well as deeming such grading to be unfair, Chinese students rated teachers in ways consistent with traditional Chinese thinking, believing that good teachers 'should impose stringent standards on students' (Ting, 2000 p. 649).

In terms of students' year level, Mason et al. (1995) could attribute no significant difference in students' ratings of teachers. Others however find that higher level subjects tend to receive higher ratings; first-year subjects, for example, were not rated as highly as those of later years, whilst ratings for graduate subjects were higher than those for undergraduate subjects (Marsh & Bailey, 1993; Neumann, 2000). However the differences tend to be small (Neumann, 2000).

Davies et al. (2007), in a study of all subjects taught in the Economics Department at The University of Melbourne, determined that some subject characteristics have a negative influence, such as whether the subject is compulsory or not and if the subject is quantitative in nature. Subjects in economics range from descriptive to highly mathematical. They also found that the relative teaching evaluations of the other subjects taken by the same cohort of students were also influential in the scores received by instructors in the same subjects.

Finally, students have been found to give lower ratings when they perceive the purpose of the evaluation process to be related to matters of staff tenure, but higher ratings when the purpose of the evaluation was seen to be the improvement of teaching in the future (Worthington, 2002).

## 2.8 Discussion

This chapter reviews part of the significant literature on the interpretation of the student surveys with particular emphasis on studies which find that the responses on the surveys are significantly influenced by factors that are not directly related to the teaching performance of the instructor, department, faculty, or institution. The range of findings in relation to student and subject level influences on quality of teaching ratings by students prompted researchers in the current study to establish the extent to which the responses on the CEQ are influenced by these factors. As has been demonstrated, the appropriate interpretation of the raw ratings may be problematic. This has led some academics to dismiss entirely any evidence gained from these types of surveys. Given the increasingly wide range of purposes for which this data is used, this seemed a dubious path to take. An analysis of how the specific factors that influence the CEQ is the topic of Chapter 5.



## Chapter 3. A systematic analysis of quality of teaching indicator surveys found in Australian tertiary institutions

This chapter presents the review of the TQIs used by the tertiary sector in Australia. We present an analysis of the question types used by these institutions and categorise these surveys by type of institution. An abbreviated version of this chapter has been published as Davies et al. (2010).

### 3.1. Introduction

Recent trends in higher education have indicated a greater emphasis on quantitative measures of institutional performance. One of the important developments in Australia has been the incorporation of data from the Course Experience Questionnaire (CEQ) (Ramsden, 1991a, 1991b; Ramsden & Entwistle, 1981) into national benchmarking and funding decisions. Since 1993, the CEQ has been conducted annually across the graduates of all universities in Australia by the Graduate Careers Council of Australia. Similar national quality assurance surveys are used in the UK, NZ and in many states of the US (Barrie and Ginns 2007).

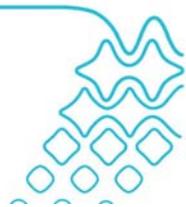
Despite the role it has assumed in the evaluation of higher education quality, the CEQ has limitations. The lagging and aggregate nature of the data make it difficult for institutions to use CEQ data alone in their internal continuous and locally-responsive quality improvement activities. The key aim of the CEQ is to measure student perceptions of their courses of study following graduation and to assess differences between academic units in terms of those perceptions (Lizzio, Wilson, & Simons, 2002; Ramsden, 1991b). The questions in the CEQ focus on student experiences, not on the characteristics of teachers or the curriculum (Ramsden 2003). The CEQ is not designed to measure student perceptions of individual lecturers or units of study.

However, in order to gain an understanding of student perceptions of individual lecturers or units of study, Australian higher education institutions have developed institution-specific instruments and surveys to provide context-relevant data and “to provide the evidence base from which to effect improvements in their performance on ... national measures” (Barrie and Ginns 2007, p. 278).<sup>2</sup> These Teaching Quality Indicators (TQIs), with names such as QOTs, SETs, LETs, TEVALs, have grown to play an important role in quality assurance in Australian higher education. In 2003 the Australian Government announced the Learning and Teaching Performance fund (LTPF) to reward higher education providers that best demonstrate excellence in teaching and learning for undergraduate domestic students. To be eligible for funding, the government requires institutions to provide evidence that the institution TQI surveys inform probation and promotion decisions for academic positions (DEST 2006).

In this study, we perform a systematic analysis of TQIs applied in tertiary institutions. By examining the nature of the surveys employed in Australian tertiary institutions we can begin to understand how different institutions monitor the quality of their instruction. In Section 3.2, we describe the data collection process. Section 3.3

---

<sup>2</sup> A unit of study is defined as a subject attracting credit points toward an award programme such as a degree.



provides an analysis of how the surveys differ across institutions and groups of institutions. Section 3.4 contains the results of an analysis of the frequency of use of questions by type. Conclusions are provided in Section 3.5.

## 3.2 The Data

We conducted an email survey of 39 Australian Universities (37 public and two private) over the period May-June 2006. In addition, we posted notices seeking information in the Higher Education Research and Development Society of Australasia (HERDSA), TeachEval and Unilearn. TeachEval is an email list of evaluation administrators. Unilearn lists academic learning advisors nation-wide. There was only one university that, at the time, did not conduct a TQI. A further two universities did not respond but they had made sufficient information available on their website. It is a requirement of all institutions requesting funding from the Australian government's LTPF that information about teaching evaluation procedures, and data derived from surveys, is publically available on university websites (DEST 2006). The data obtained from email responses and websites can therefore be considered accurate and reliable. In order to ensure that the survey data we had obtained through email and web site data was as accurate and reliable as possible, we established an online database called "Evaluation Central" in which institutions can themselves confirm the accuracy of the data as well as the way in which items in the survey for their particular institution have been classified according to the classification system of question types that we devised<sup>3</sup>.

### 3.2.1 The Institutional Practices

We surveyed institutions to obtain information on: the name of their survey; whether there was a separate survey for units of study and lecturers; whether evaluation was conducted online, in paper-based form, or both; the period of data collection; whether the TQI was compulsory, effectively mandatory or optional<sup>4</sup>; whether the data was available for research purposes; the name of the unit responsible for collecting the data; the number of core questions in the survey and whether there were open response questions.

### 3.2.2 The Question Types

Whilst it is generally considered that students' perceptions of institutional quality are multidimensional in nature there is less agreement as to the number and nature of the dimensions (Jackson et al. 1999). We classified the questions within the surveys according to the type of information that was sought by each institution. We devised a schema which clustered similar questions together by developing our own classification of 18 question "types" (QTs). For example, a number of institutions ask for responses to a question that is similar or identical to: "*This subject is well taught*" (Question 2 in The University of Melbourne survey). In our schema, all questions of this nature were clustered under the question type: "*Overall Teaching Quality*". Other questions were similarly grouped under question types capturing the perceived intent of the question.

---

<sup>3</sup> The complete list of the surveys from each institution is provided in Appendix A.

<sup>4</sup> "Compulsory" was defined as being a systematic, institution-wide practice required by the university and conducted on a regular or semi-regular basis for all teaching staff. "Effectively mandatory" was defined as being necessary for promotion or advancement purposes.



Several discussions were undertaken with the team in order to categorize the questions. We established an online database called Evaluation Central<sup>5</sup> whereby Evaluation Managers from each institution were invited to register and to confirm or amend the question typology that was proposed.

There were two distinct groupings of questions:

- 1) about the lecturer and the subject; and
- 2) about the student and their learning.

The details of each Question Type are given below.

### **Question Types relating to the Lecturer and Subject**

#### **1. Clear Aims:**

This refers to the clarity of the *aims of the class or subject or course* in terms of standards and objectives, not the clarity of the lecturer or the teaching (the latter is captured under “Clear Explanations”). For example: “The subject objectives were made clear to me”.

#### **2. Clear Explanations:**

This captures the clarity of the lecturer in giving explanations either a) in general terms, or b) in outlining expectations of the course. For example: “The lecturer was able to communicate concepts clearly”.

#### **3. Organised:**

This refers to the extent to which either the lecturer or the subject or unit was well-organised, well-prepared and well-structured. For example: “The teaching of this unit is well-organised”.

#### **4. Motivation/Enthusiasm of Lecturer:**

This refers to the level of teacher’s enthusiasm in teaching. For example: “The lecturer was enthusiastic about the subject”.

#### **5. Respect:**

This refers to the lecturer’s sensitivity to students’ problems, politeness and friendliness to students and their cultural backgrounds and/or their different views and opinions. For example: “The lecturer was sensitive to students’ cultural backgrounds”.

<sup>5</sup> The address for this website is <http://tlu.ecom.unimelb.edu.au/evaluationcentral>



6. *Access:*

This refers to the extent to which lecturers were available for consultation outside normal lecturing times. For example: “The lecturer was available to answer students’ inquiries”.

7. *Teacher knowledge:*

This refers to the *perceived understanding by students* of the lecturers’ knowledge of the content/subject matter that he or she was teaching. For example: “The lecturer has a sound knowledge of the topic”.

8. *Overall Teaching Quality:*

This captures the overall teaching quality of the lecturer. For example: “This teacher communicates effectively with students/This subject is well-taught”.

**Question Types relating to the Student and their Learning**

9. *Motivation/Enthusiasm of Student:*

This refers to the level of student motivation and enthusiasm. For example: “I am motivated to achieve learning outcomes”.

10. *Student knowledge:*

This refers to whether the students felt that—as a result of the lecturer’s classes—they had gained an understanding of the subject matter. For example: “In this teacher’s class I have gained a good understanding of the concepts covered”.

11. *Stimulating/Interesting/Motivating:*

This refers to the level of interest generated on the part of the student from the classes. Did the lecturer inspire the students, motivate them, get them to think, challenge them? For example: “The teaching staff motivated me to do my best work”.

12. *Gave Feedback:*

This refers to whether the lecturer made time to assist students with the learning needs and problems. For example: “The feedback on my work is provided promptly”.

13. *Assessment:*

This refers to the nature and effectiveness and clarity of the assessment tasks requested by lecturers in assessing students’ understanding of the subject content. For example: “Overall the assessment in this unit is fair”.



**14. Students' Needs and Learning Skills:**

This refers to whether lecturers were sensitive to students' learning needs and to the extent to which the lecturer actively developed learning skills (critical thinking, discursive knowledge, understanding rather than memorizing, etc). For example: "My learning in this subject was well supported".

**15. Receives Feedback:**

This captures the extent to which student feedback was encouraged and whether the feedback was used to improve teaching. For example: "The teacher shows genuine interest in improving his/her teaching".

**16. Teaching Methods/Material/Aids Used:**

This refers to the students' perception of teaching aids and methods used for teaching. Were they useful, effective, and/or relevant? For example: "The teacher related the course materials to real life situations". "I found the teaching methods used in this subject were effective in helping me to learn".

**17. Workload:**

This refers to the workload expected. Was it commensurate with expectations, fair or unreasonable? For example: "The workload was appropriate for a subject at this level".

**18. Overall Effectiveness:**

This is an overall judgment by the students on the lecturer's effectiveness and/or the effectiveness of the unit or subject taught. For example: "Overall how would you rate the learning experience in this course".

**Miscellaneous Question Type:**

This category is for questions, open comments, etc., that do not naturally fit the other categories. For example: "Work marked by this teacher is returned within a reasonable time".

The classifications of the TQIs from each university in this study are provided in Table B.1 in Appendix B in the Resource Document for this report.

### 3.3 A Descriptive Analysis of Surveys Used by Institutions

In 1998 the Graduate Careers Council of Australian (GCCA 1999) undertook the *Institutional Arrangements for Student Feedback (IASF)* project. One of the purposes of this project was to prepare an inventory of instruments used by universities to survey students and to prepare an analysis of common elements. It was concluded that: "in the teaching and learning area, for example, often they are the same surveys or are modifications from the same two or three originals ..." (p. 14, GCCA 1999). However, Barrie and Ginns (2007, p. 278) claim that "each university develops its surveys in isolation". Here we investigate more precisely the



interrelationship between these surveys.

In this section, we examine the surveys to determine how their form may be influenced by the institution that uses them. This is based on two analyses. The first uses the proportion of all question types that are student-oriented to determine if there are patterns by the group to which the university belongs. In the second part of this section we report on the results of a cluster analysis of institutions by type of survey they use.

### 3.3.1 The Classification by Focus on Student Learning

The schema developed in Section 3.2 contains two distinct groupings defined by questions about the lecturer and subject and questions about the student and learning. For the TQI of each institution, the student learning quotient (SLQ) is defined as:

$$SLQ = \left( \frac{\text{number of student and learning type questions}}{\text{number of total question types used}} \right) \times 100$$

These are reported in Table 3.1 by the groupings of each university and the institutions within each grouping are ordered by their SLQ (as of 2006 when the questionnaires were collected). The average across all universities in our set is SLQ =55.8 which indicates that there is a slightly higher proportion of student and learning QTs.

Institution	SLQ	Institution
<b>Group of Eight (GO8)</b>		<b>New Generation Universities (NGU)</b>
The University of Melbourne	71	Victoria University
The University of New South Wales	67	University of Ballarat
The University of Sydney	67	CQUniversity Australia
The Australian National University	57	Edith Cowan University
The University of Adelaide	50	University of Western Sydney
The University of Western Australia	44	University of the Sunshine Coast
The University of Queensland	40	University of Canberra
Monash University	36	University of Southern Queensland
<b>Innovative Research Universities (IRU)</b>		<b>Southern Cross University</b>
Griffith University	78	<b>Independents (IND)</b>
La Trobe University	50	University of Wollongong
Macquarie University	50	Deakin University
The University of Newcastle	50	Charles Sturt University
Flinders University	40	Swinburne University of Technology
Murdoch University	36	Charles Darwin University
<b>Australian Technology Network (ATN)</b>		James Cook University
Curtin University of Technology	78	University of New England
Queensland University of Technology	78	University of Tasmania
University of South Australia	63	
University of Technology, Sydney	63	
RMIT University	57	

**Table 3.1: A Comparison of SLQ by Institution**

In Table 3.1, we consider five groupings of Australian Universities defined as: the



Group of Eight (GO8), the Australian Technology Network (ATN), the Innovative Research Universities (IRU), the New Generation Universities (NGU) and we refer to the remaining institutions as Independents (*IND*). From Table 3.1 we note that there is a range of SLQ values across each grouping of institutions. . The university with the greatest ratio of student and learning questions in their TQI is the University of Wollongong, whereas the University of Tasmania and Southern Cross University both have the lowest ratio which shows greater attention to lecturer and subject question types. Interestingly all of the ATN institutions have surveys with a higher SLQ than average, while, all of the IRU institutions have SLQs less than average except for Griffith University. This suggests that the technology- oriented institutions give more emphasis to student and learning question types, whilst the Innovative Research Universities focus their surveys more on lecturer and subject characteristics. Other groupings of institutions demonstrate no particular pattern in their SLQs.

### 3.3.2 A Cluster Analysis of the Universities

We applied an agglomeration cluster analysis where the similarity measure is defined as the number of questions of the same type that each university's survey used based on the schema developed in Section 3.2 (See Russell & Rao, 1940).<sup>6</sup> Using the furthest-neighbour measure between clusters, we clustered the universities into groups described by the dendrogram in Figure 3.1.<sup>7</sup> In the far left of the figure each institution is in a cluster with only one member. The closest universities are combined first, then the next, progressively until there is only one cluster. The dissimilarity between the members of the cluster is given by the length of the horizontal line to the point where they join.

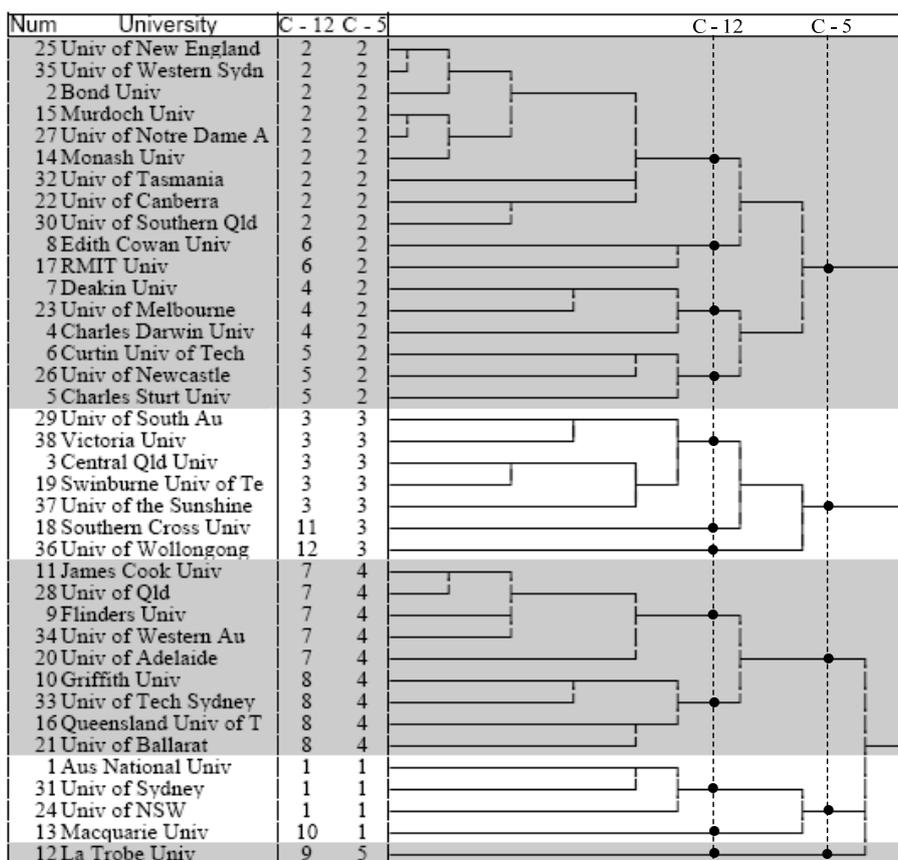
The clustering process can be stopped once a specific number of clusters have been formed. For example, the membership of those clusters defined when there are 5 clusters is given in Figure 3.1 where each cluster group is shaded differently and identified in Column C-5. From Figure 3.1 we note that, for the case of 5 clusters, the largest group of similar surveys is composed of 17 universities. Aside from defining groups, cluster analysis can also identify "outliers": in this case we find that La Trobe University alone defines one of the five clusters. From Figure 3.1 we can also identify the membership in the 12 clusters as shown by column C-12. Those universities identified as "outliers" in Figure 3.1 might investigate whether they should include information that other universities are using in their measurement of teaching effectiveness.

From Figure 3.1 we can determine if the cluster definitions coincide with the groupings of Australian institutions as defined in Table 3.1. Three GO8 universities (ANU, Sydney, and NSW) are all included in cluster C-12,1, with three additional GO8s (Queensland, Western Australia, and Adelaide) in cluster C-12,7. Only two Victorian institutions (Monash and Melbourne) are in separate clusters. We also note that four members of the NGU (Canberra, Southern Queensland, Western Sydney and Edith Cowan) are clustered in C-5,3. Of the NGU, only Ballarat is in a different cluster. Thus we observe a correspondence between the group definitions for GO8 and NGU institutions and the clusters to which they are allocated.

<sup>6</sup> Table B.2 in Appendix B of the Resource Document for this report presents the data and the proximity matrix between the various universities' TQIs used for this cluster analysis.

<sup>7</sup> More detail as to the process by which the agglomeration cluster method operates can be found in Kaufman and Rousseeuw (1990).





**Figure 3.1: The Dendrogram of the hierarchical cluster analysis of Australian universities created using a complete linkage method. The C-5 and C-12 columns indicate the membership in either the 5 group clustering or the 12 group clustering.**

In order to identify the characteristics of each cluster we constructed a typical survey by cluster in Table 3.2. This was done by noting which QTs had more than 50% of the cluster members use them in their survey with a marker (●). From Table 3.2 we note that the largest cluster in the five cluster set (#2) has the largest number of QTs and that most of the smaller clusters have surveys that use fewer QTs. We also note that there are some QTs that are employed much more widely than others. In the next section we investigate the relationship between the QTs used on each survey as defined by the schema developed in Section 3.2 and their frequency of use.



Question Types (QTs)	12 Clusters												5 Clusters				
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5
1 Clear Aims		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
2 Clear Explanations	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
3 Organised	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
4 Motivation/Enthusiasm of Lecturer							•	•	•	•	•	•				•	•
5 Respect		•												•			
6 Access		•				•	•	•						•			
7 Teacher Knowledge											•						
8 Overall Teaching Quality		•	•	•	•	•	•	•	•	•	•	•		•			
9 Motivation/Enthusiasm of Student												•					
10 Student Knowledge												•					
11 Stimulating/Interesting/Motivating	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
12 Gave Feedback	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
13 Assessment												•					
14 Students Needs and learning Skills	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
15 Receives Feedback												•					
16 Teaching Method/Material/Aids used												•					
17 Workload												•					
18 Overall effectiveness	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

**Table 3.2:** Typical Surveys by Cluster Sets membership in the 5 and 12 cluster cases.

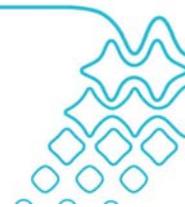
### 3.4. An Analysis of Questions by Type.

In this section we first present the differences by the institutions’ membership in different groups and then we show a cluster analysis to establish the similarity between the patterns of question types used across institutions.

#### 3.4.1 Question Type Usage by Institutional Grouping

By considering the five groupings of Australian universities defined in Table 3.1, we examined the tendency for certain groups to employ particular QTs more than others. Because membership in these groups is based on similarities in scale and focus we might expect that this would influence the types of information they request.

Table 3.3 reports the proportion of each group that uses each QT. The cells in Table 3.3 are ranked by the frequency of QT use by *All Institutions*. Both the QTs most likely (12, 18 and 14) and least likely (17, 10 and 15) to be used fall into our category of QTs about the students and their learning. QTs 1 and 2 are the most likely questions to be asked that fall into our category of QTs about the lecturer and subject. There are also some distinct patterns of QTs that seem to follow institutional patterns. Within the groupings of the universities, all members of the *G08* ask QTs 14 and 11, all members of the *ATN* ask QTs 12 and 18 and all members of the *NGU* ask QT 18. However, there is no one question that is asked by all members of the *IRU*. In addition, many of the members of the *ATN* ask QTs 9 and 16 which are both QTs that are not commonly asked by *All Institutions*. In comparison, the *G08* members are unlikely to pose QT 1 which is the 5<sup>th</sup> ranked question by *All Institutions* and none of the *G08* use QT 13 which is the 9<sup>th</sup> ranked question by *All Institutions*. Another interesting aspect of the use of different question types is that those members of the Independent group, as well as the



members of the *ATN*, are almost twice as likely to include a questions relating to teaching methods and materials (QT 16) than others.

<b>Question type (QT)</b>		<b>G08</b>	<b>ATN</b>	<b>IRU</b>	<b>NGU</b>	<b>IND</b>	<b>All</b>
12	Gave Feedback	75	100	83	78	88	84
18	Overall effectiveness	88	100	67	100	63	82
14	Students Needs and learning Skills	100	60	67	78	75	79
2	Clear Explanations	88	60	83	78	38	71
1	Clear Aims	38	80	50	89	63	66
11	Stimulating/Interesting/Motivating	100	60	33	56	75	66
3	Organised	75	20	67	56	50	58
8	Overall Teaching Quality	38	40	33	44	75	50
13	Assessment	0	60	50	56	63	47
6	Access	38	20	50	44	50	45
4	Motivation/Enthusiasm of Lecturer	50	20	50	44	25	40
7	Teacher Knowledge	25	20	17	56	50	37
9	Motivation/Enthusiasm of Student	25	80	17	33	38	34
16	Teaching Method/Material/Aids used	25	60	17	22	63	34
5	Respect	38	0	33	22	25	26
17	Workload	13	20	17	11	13	16
10	Student Knowledge	0	0	17	11	25	11
15	Receives Feedback	0	20	17	33	0	11

**Table 3.3: The percentage of each group that uses each QT, ranked by frequency of all institutions use of the questions.**

Some of these results are also evident from Table 3.2 which reports the typical survey by cluster in order to identify the characteristics of each cluster. From this Table 3.3 we find that QTs 4 to 10 and 15 to 17 are not widely employed whilst QTs 2, 12, 14, and 18 are more widely used. Also we note that the single institutions that comprise clusters 9 to 12 pose few of the QTs from 5 to 11 and 15 to 17.

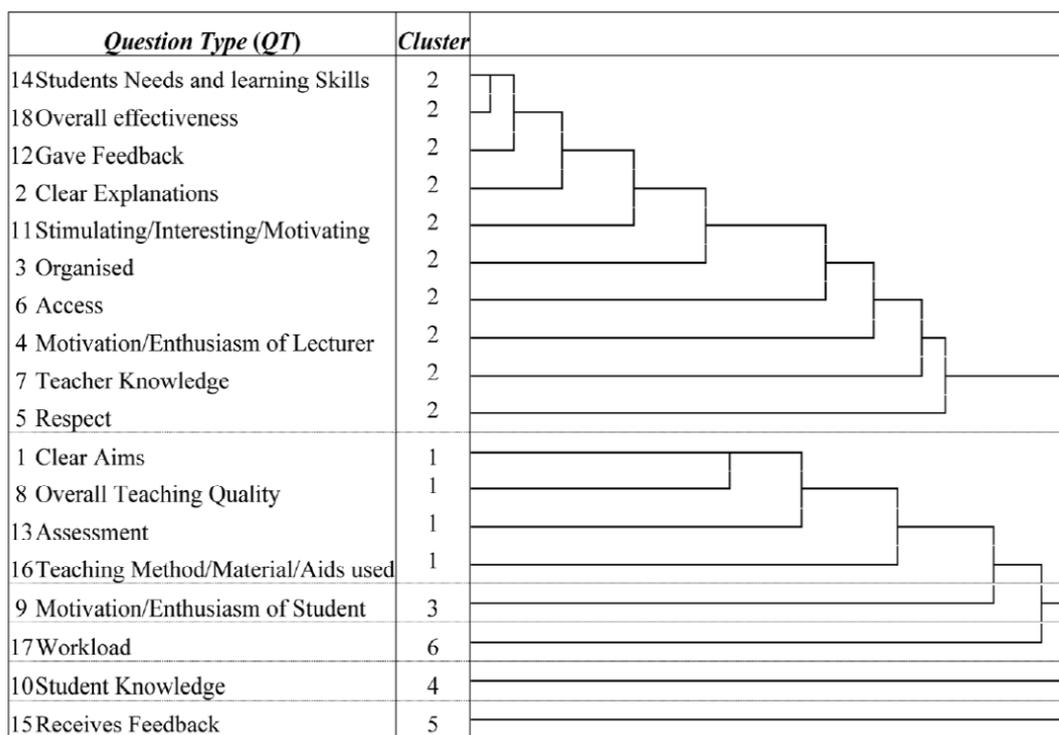
### 3.4.2 A Cluster Analysis of the Question Types

In order to determine the relationship between the various questions asked in the surveys, we used a cluster analysis based on frequency that these QTs are employed. A similarity matrix to compare QTs is defined by determining the number of universities that pose each type of question.<sup>8</sup>

Using the 18 by 18 similarity matrix of counts of universities that use the same QTs, we apply an agglomeration cluster analysis where the inter-cluster similarity is measured using the complete linkage distance. From Figure 3.2 we note that if we stop the clustering algorithm when there are six clusters QTs 9 (Motivation/Enthusiasm of Student), 17 (Workload), 10 (Student Knowledge), and 15 (Receives Feedback) are placed in their own cluster. QTs 17, 10 and 15 are the most rarely posed question types. However QT 9 is used by more than a third of the institutions of which the predominant portion consists of members of the *ATN*. It is also noteworthy that QTs 1, 8 and 13, although widely used as noted from Table 2, are not included in the largest cluster.

<sup>8</sup> Table B.3 in Appendix B in the Resource Document for this report lists the entire table of proximities by question.





**Figure 3.2: The Dendrogram of the Hierarchical Cluster Analysis of the QTs based on a complete (or furthest neighbour) linkage method.**

### 3.5. Conclusions

National quality assurance surveys that gather data from graduates on their experiences of their entire course of study are used in a number of countries including Australia. However, to gain an understanding of student perceptions of individual lecturers and units of study institution-specific surveys are often used. In this chapter we have examined the TQIs used by Australian universities. We have categorized the questions used in these surveys into question types in order to determine if questionnaire form can be used to identify how different universities approach the measurement of teaching effectiveness by their academic staff.

We explored how different universities were similar to each other based on which questions were used. In order to establish these groupings we employed a cluster analysis to provide groupings of universities based on the types of questions they include in their TQIs. The result of this analysis found that some universities use a TQI that is quite distinct from the majority of other institutions. This is important information for universities and could inform any redesign of their internal TQI. If particular institutions wish to benchmark themselves against those institutions with which they wish to be aligned, a review of the questions they use in the TQI would appear to be in order. We found that all members of the GO8 ask question types 14 and 11, all members of the ATN ask question types 12 and 18 and all members of the NGU ask question type 18. We also found that the ATN institutions ask a higher percentage of student learning questions than other institutional groups and that practically all the IRU institutions ask fewer than average. From our cluster analysis, we observe a correspondence between the group definitions for GO8 and NGU institutions and the clusters to which they are allocated.



Besides the analysis QTs by university group , we also investigated any patterns in their use. To accomplish this we defined a matrix of similarity between the question types as defined by our classification and based on the number of universities that asked the same questions. The outcome of this analysis revealed that questions relating to increases in student knowledge; the degree to which the student's feedback to the instructor is encouraged or not; and the expectation of workload (10, 15, and 17) were not commonly included in the types of questions asked. Whereas, questions concerning Feedback, General Effectiveness, and Student's Needs and Learning Skills (12, 18, and 14) were far more likely to be included in the survey.

Even though a form of TQI is used in almost all Australian universities, we find that the surveys vary from institution to institution. Based on the survey results and the analysis conducted in this chapter, it is anticipated that Australian institutions can be better informed as to how their TQIs compare to those conducted by other tertiary institutions and can consider adjusting their TQI based on the analysis conducted here.



## Chapter 4. The matching of quality of teaching indicators and the course experience questionnaire<sup>9</sup>

In this chapter we present an analysis of the responses to a set of questionnaires based on the CEQ, the TQI from a number of different Universities and a new TQI based on the types of questions defined in our analysis in Chapter 3. By conducting surveys of students in a number of different universities, we have attempted to establish the degree to which the TQI from these institutions and the new TQI based on the research reported in Chapter 3 correspond to the responses to the CEQ questions that are part of the Good Teaching Scale (GTS), the Generic Skills Scale (GSS) and the Overall Satisfaction Item (OSI). This chapter proceeds as follows: first, we introduce the problem and form of our examination. In Section 4.2 and 4.3 we describe the universities, the subjects used in our analysis and the CEQ and the new TQI based on our research in Chapter 3. In Section 4.4 we describe the results of a number of different analysis techniques that were designed to determine the interrelationship between the responses to the new TQI and the CEQ questions. This analysis was conducted using the Rasch Item Response Methodology, Cluster Analysis, Factor Analysis and Regression Analysis. In Section 4.5, we use the Rasch Item Response Methodology to demonstrate the relationship between the local TQI and the CEQ.

### 4.1. Introduction

The Australian government surveys all graduates of all nationally funded tertiary Institutions using a common survey instrument. This survey is called the Course Experience Questionnaire (CEQ) (Ramsden, 1991a, 1991b; Ramsden & Entwistle, 1981) and is designed to gather information on a number of aspects of students' experiences during their course of study at the institution. Some of the results of these surveys have been used to determine the funding of special grants for institutions that score well (*Learning Teaching and Performance Fund*) and have also been used to make comparisons across universities in widely available guides such as the *Good Universities Guide* (Evered 2008).

The CEQ was developed for two main reasons. The first was to provide a performance indicator of teaching effectiveness in higher education institutions; the second, to enable institutions with their quality enhancement and improvement processes (see eg. Griffin et al. 2003). However, a well-known limitation of the CEQ is its lagging nature. Usually, the final reports of graduating students' experiences are not available until at least a year after they finish their courses. Consequently, this makes it difficult for universities to use the CEQ data alone in their internal continuous and locally-responsive quality improvement activities. The CEQ is not designed to measure student perceptions of individual lecturers or units of study. Institution-specific instruments and surveys have been designed to gain an understanding of these perceptions. Most institutions in Australia conduct end-of-semester surveys or TQI within each subject (or class) taught.

---

<sup>9</sup> Earlier versions of this chapter were presented at The 8<sup>th</sup> Annual Hawaii International Conference on Education, Honolulu, HI, January 10, 2010, The Forum on Quantitative Analysis Of Teaching And Learning In Higher Education In Business, Economics And Commerce, The University of Melbourne, February 12, 2010, and The 15<sup>th</sup> Australasian Teaching Economics Conference, Hamilton New Zealand, June 29, 2010.



In this analysis, we surveyed students in four different institutions to establish the relationship between the TQI responses on the locally administered survey and the responses on the CEQ survey in order to establish the degree to which the indications of good teaching as defined in the CEQ coincide with the positive responses elicited from the TQIs. This is done using a series of questionnaires that include 3 sets of questions: those from the local TQI, those from a TQI that is the same for all institutions based on the types of questions we describe in Chapter 3, and a subset of the questions from the CEQ that relate to the quality of teaching.

This section proceeds as follows. Section 4.2 describes the four universities and subjects that were surveyed. Section 4.3 provides details of the constructed TQI and the questions that were used from the CEQ that made up the surveys. Section 4.4 presents the results of four different analyses that were used to establish the degree to which the indications of good teaching, as defined in the CEQ, coincide with the positive responses elicited from the TQI. Section 4.5 presents conclusions.

## 4.2. Description of Universities

Surveys based on the CEQ and TQI (and explained in detail in the next section) were administered to a number of students in four separate Universities. The institutions where these surveys were administered are:

*University 1:* Flinders University in Adelaide, South Australia, medium sized and established in the 1960's.

*University 2:* The University of Melbourne in Melbourne, Victoria, a large city based university including a large international student population that was established in the 1850's

*University 3:* The University of Tasmania in Hobart, Tasmania, a small university established in the 1890s.

*University 4:* The University of Wollongong in New South Wales, a large university with a large international student population including international campuses, with links to the regional community, and established in the 1970s.

Using data from the Course Experience Questionnaire and adjusting for characteristics of the universities, Williams (2008) constructed a ranking of Australia's 37 public universities based on student satisfaction in undergraduate degrees. The rankings of the four universities used here fall within the top 10%, 30%, 65% and 80% range. This mix of institutions covers a range from primarily teaching (Flinders and Wollongong) to those that provide full offerings at both the undergraduate and graduate level (Tasmania) and one that is rated as a world class research institution (Melbourne).

All subjects used were from the Commerce Faculty. A small subject from the Commerce Faculty was surveyed from University 1. The subjects surveyed in University 2 included a large compulsory second year subject from the Management Department, that was taught in a number of sessions and a third year subject from the Finance Department. A number of small subjects were surveyed in University 3 from the School of Management and the School of Accounting and Corporate Governance, including both second and third year subjects. A large second year subject was surveyed from the Management Department from the main campus for



University 4.

### 4.3. Description of Surveys

In this study we examine the relationship between a constructed TQI and the CEQ. Each university uses an institution-specific survey to gain an understanding of student perceptions of individual lecturers and units of study. We use the categorization of questions developed in Chapter 3 (Davies et al., 2010) to develop a generic TQI.

#### 4.3.1 Australian TQI Surveys as described in Chapter 3.

In Chapter 3 we examined the TQIs from all Australian universities and classified the questions to explore how different universities' surveys are similar to each other. Based on the questions asked by the separate universities, a set of questions were devised to generalise these specific questionnaires into a signal set of survey questions. The questions in this generic survey are presented in Table 4.1.

N1	I had a clear idea of what was expected of me in this subject
N2	The lecturer was good at explaining the subject matter
N3	The lecturer was well organised
N4	The lecturer communicated enthusiasm for the subject area
N5	The lecturer treated students with respect
N6	The lecturer has been available to discuss problems and questions relating to my assignments & exams
N7	The lecturer knew the subject matter well
N8	This subject was intellectually stimulating
N9	The lecturer assisted me in gaining a good understanding of the subject matter
N10	The lecturer motivated me to learn
N11	I received helpful feedback on how I was going in this subject
N12	The lecturer clearly explained what I was required to do in assessment items
N13	Lecturers showed an interest in the academic needs of the students
N14	There was effective use of computer-based teaching materials in this subject
N15	The lecturer presented an appropriate amount of material for the time available

**Table 4.1: New TQI administered at four universities**

#### 4.3.2 CEQ

The CEQ survey gathers data on students' perceptions of their course using 25 items, which aggregate to five factors and a single item. In this study we examined 3 of the five factors in the CEQ that are directly related to subject or class characteristics: the Good Teaching questions and those related to Generic Skills. We also include the single Overall Satisfaction questions.

C1:	The lecturers put a lot of time into commenting on my work.
C2:	The lecturers normally gave me helpful feedback on how I was going.
C3:	The lecturers of this course motivated me to do my best work.
C4:	My lecturers were extremely good at explaining things.
C5:	The lecturers worked hard to make their subjects interesting
C6:	The lecturers made a real effort to understand difficulties I might be having with my work.

**Table 4.2: The Good Teaching Scale**



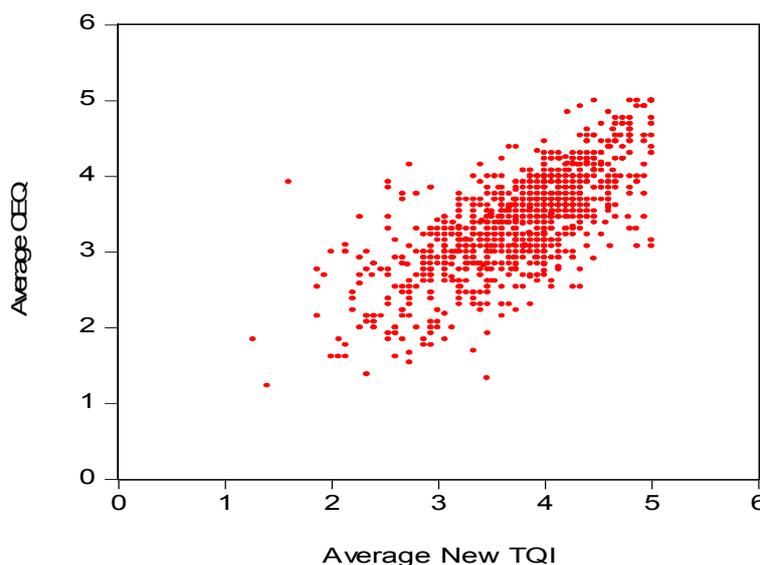
C7	The course helped me develop my ability to work as a team member.
C8	The course sharpened my analytic skills.
C9	The course improved my skills in written communication.
C10	The course developed my problem-solving skills.
C11	As a result of my course, I feel confident about tackling unfamiliar problems.
C12	My course helped me to develop the ability to plan my own work.

**Table 4.3: The Generic Skills Scale**

C13	Overall, I was satisfied with the quality of this course.
-----	-----------------------------------------------------------

**Table 4.4: The Overall Satisfaction Item**

The *Good Teaching Scale* (GTS) is characterised by practices such as providing students with feedback on their progress, explaining things, making the course interesting, motivating students, and understanding students' problems. The six items that comprise the GTS are listed in Table 4.2. The *Generic Skills Scale* (GSS) is an attempt to take into account the extent to which university courses add to the generic skills that their graduates might be expected to possess. The six items comprising the GSS scale are presented in Table 4.3. The single overall satisfaction question as given by The Overall Satisfaction Item (OSI) is presented in Table 4.4.



**Figure 4.1:** The average score for the new TQI and CEQ questions.

#### 4.4 Results of analysis of New TQI and the CEQ

Three techniques were used to assess to what extent the TQIs align with the CEQ. These techniques included Item Response Modelling, factor analysis and cluster analysis. A preliminary view of the data is shown in Figure 4.1 where the average response to the new TQI (N1-N15) and the average response to the CEQ (C1-C13) are plotted. From this plot we can see that there is an overall positive relationship between the averages, however the agreement is not always complete and the bounded nature of these measures provides an unrealistic impression of the

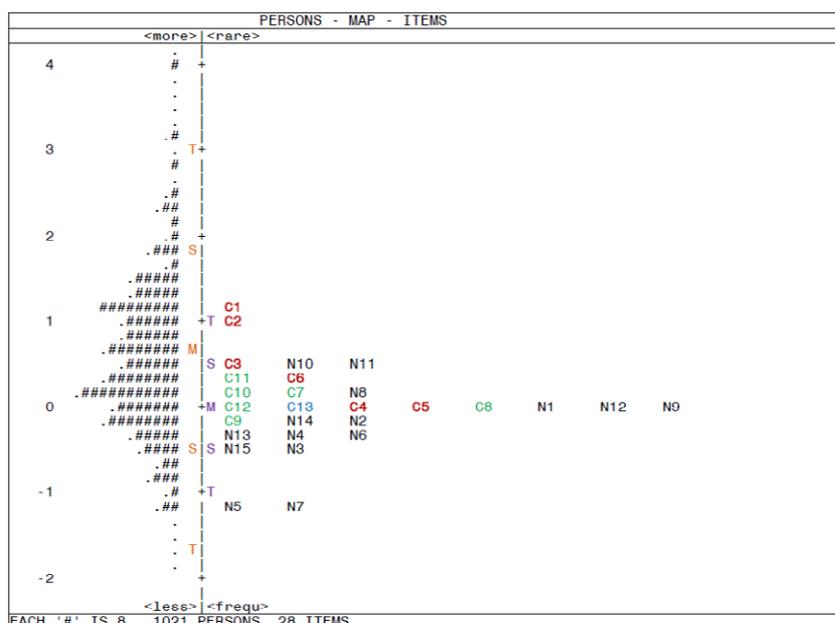


relationship between them. The correlation between these averages is 0.70. In this section a number of techniques are demonstrated which were employed to establish the interrelationship between the individual items in the CEQ and the new TQI.<sup>10</sup>

#### 4.4.1 Item Response Modelling

The first method of analysis applied to compare the responses to these questions was the Item Response Modelling method or what is commonly referred to as the Rasch Model (RM) (see for example Bond and Fox, 2007). The RM is widely applied in the analysis of testing methods and provides a technique for the determination of the degree to which test results are due to the questions asked as well as the ability of the test takers.

In this case it is used primarily to determine the degree of “difficulty” in responding positively to some questions versus others. The survey questions are posed as positive statements and the agreement or disagreement with the statement is measured on a Likert scale. The total “score” from each student surveyed can be thought of as equivalent to the total marks scored on a test as all the questions are positive. In the same manner the degree of “difficulty” can be thought of as a degree to which the question leads to positive responses.



**Figure 4.2:** Item Person Map for new TQI and CEQ for four universities

Figure 4.2 presents the Item person map output from the RM. The RM shows that many of the items on both the CEQ and the new TQI elicit responses have similar levels of response by the surveyed students. However they are not close substitutes for the same level of “positiveness”. From the Item person map we note that questions C2 (*helpful feedback*), C1 (*comments on work*), and C3 (*motivated me*) are grouped together as harder to score highly, whilst the new TQI questions are located where it is easier to get higher values.

<sup>10</sup> The average responses by item and institution are listed in Appendix C of the Resource Document for this report in Table C.1.



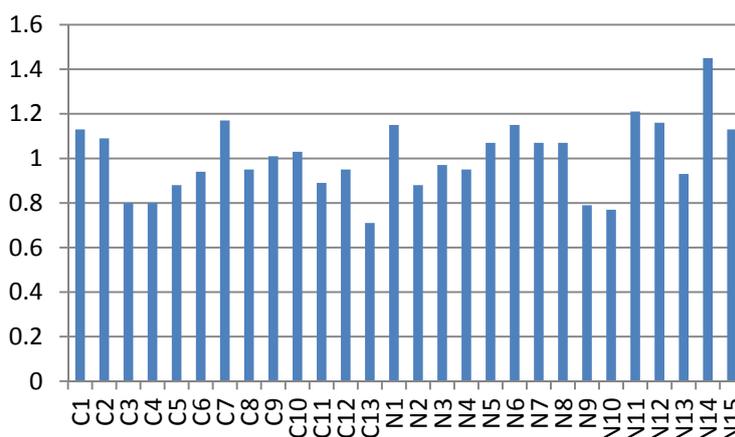


Figure 4.3: Fit (mean square): CEQ and New TQI items

Fit	Item	Item
1.45	N14	There was effective use of computer-based teaching materials in this subject
1.21	N11	I received helpful feedback on how I was going in this subject
1.15	N6	The lecturer has been available to discuss problems & questions relating to my assignments or exams
1.15	N1	I had a clear idea of what was expected of me in this subject
1.16	N12	The lecturer clearly explained what I was required to do in assessment items
1.17	C7	The course helped me develop my ability to work as a team member
1.13	C1	The lecturers put a lot of time into commenting on my work
1.13	N15	The lecturer presented an appropriate amount of material for the time available
1.09	C2	The lecturers normally gave me helpful feedback on how I was going
1.07	N8	This subject was intellectually stimulating
1.07	N5	The lecturer treated students with respect
1.07	N7	The lecturer knew the subject matter well
1.03	C10	The course improved my skills in written communication
1.01	C9	The course developed my problem-solving skills
0.95	N4	The lecturer communicated enthusiasm for the subject area
0.97	N3	The lecturer was well organised
0.95	C8	The course sharpened my analytic skills
0.94	C6	The lecturers made a real effort to understand difficulties I might be having with my work
0.95	C12	My course helped me to develop the ability to plan my own work
0.93	N13	Lecturers showed an interest in the academic needs of the students
0.88	N2	The lecturer was good at explaining the subject matter
0.89	C11	As a result of my course, I feel confident about tackling unfamiliar problems
0.88	C5	The lecturers worked hard to make their subjects interesting
0.8	C3	The lecturers of this course motivated me to do my best work
0.77	N10	The lecturer motivated me to learn
0.8	C4	My lecturers were extremely good at explaining things
0.79	N9	The lecturer assisted me in gaining a good understanding of the subject matter
0.71	C13	Overall, I was satisfied with the quality of this course

Table 4.5: Fit statistics for the new TQIs versus CEQ.

The Fit statistics are presented in Table 4.5 and plotted in Figure 4.3. These indicate that C13 (*overall satisfaction*) question for the CEQ and N9 (*helped to gain a good understanding*) question for the New TQI and C4 (*the lecturers are good at explanations*) received a a much lower proportion of responses than would be predicted by the responses in general. The questions that elicited the over fit values are the ones that relate to the New TQI in particular: N14 (*effective computer use*), N11 (*effective feedback*), N6 (*availability of lectures to discuss exams*), N1 (*clear idea of goals in subject*) and N12 (*clear definition of assessment tasks*).

#### 4.4.2 Factor Analysis

An alternative analysis to consider how these survey questions compare with each other is Factor Analysis (see Morrison 1976). Where the RM uses the logistic



transformation of the scores, here we use the actual values in a correlation matrix. The Factor Analysis here is based on the responses from all four institutions for the relationship between the CEQ and the new TQI proposed for all institutions.

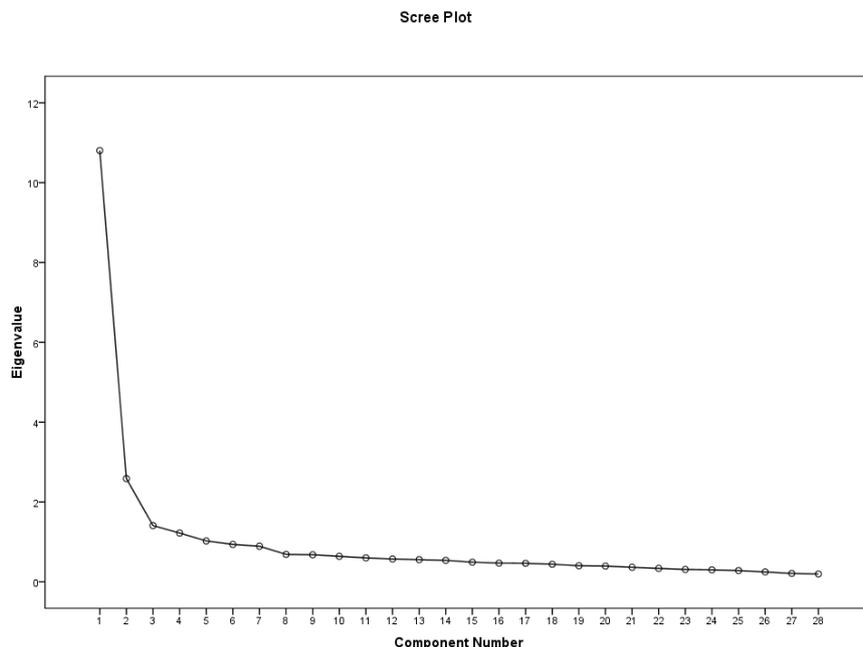


Figure 4.4: Scree Plot of Eigenvalues by number

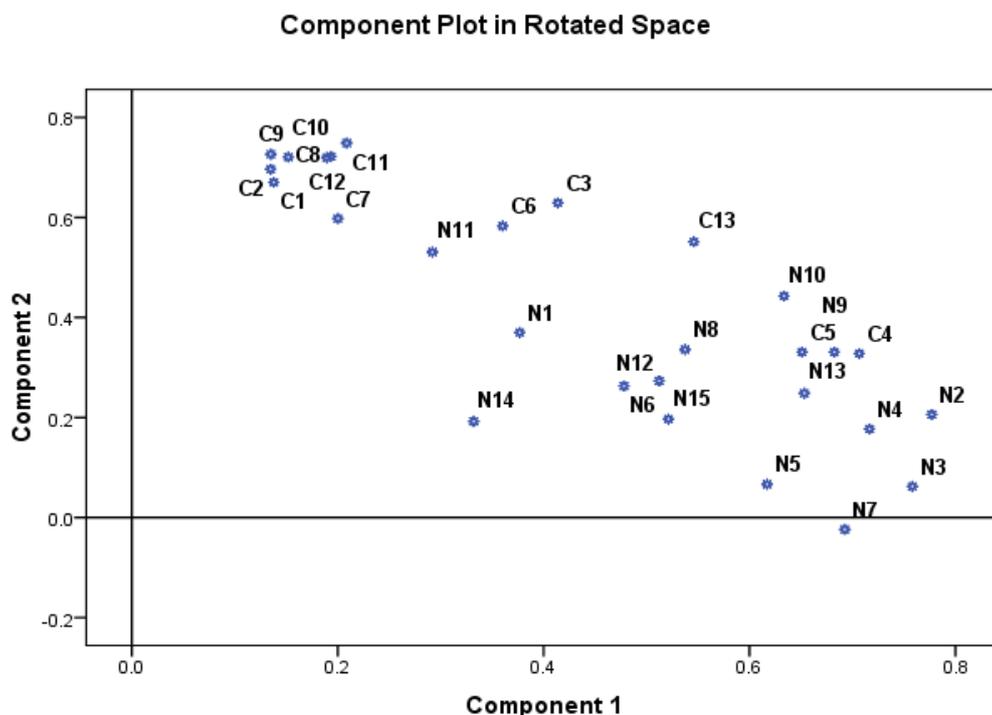


Figure 4.5: The Component Plot of the first two Factors that explain the most variation in the data (Using a varimax rotation)

In this analysis we found that the first two principle components accounted for 48%



of the variation in the data. The scree plot in Figure 4.4 indicates that, once we consider the first two components, the rest of the dimensions account for much less of the variation in the data. The scree plot indicates that most of the variation in the correlation matrix of the question response can be found from the first two eigenvalues.

Figure 4.5 is a plot of the first two components. We found that for the most part the two types of questions (N and C) receive distinctly different responses. Most of the CEQ questions, with the exception of C5 (*The lecturers worked hard to make their subjects interesting*) and C4 (*My lecturers were extremely good at explaining things*) can be grouped apart from the new TQI questions with the exception to N11 (*I received helpful feedback on how I was going in this subject*).

This idea that the questions may be grouped or “clustered” by response patterns lends itself to an alternative way of thinking about how the students responded to these questions. The subsequent cluster analysis was based on a measure of similarity that was designed to consider Likert scale results.

#### 4.4.3 Cluster Analysis

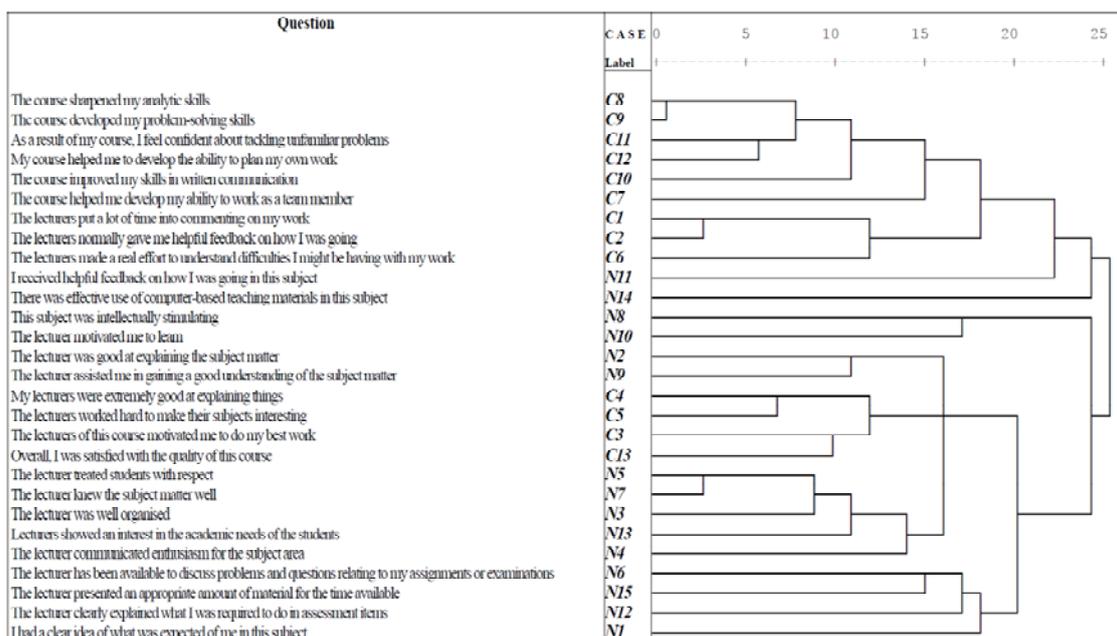
Cluster Analysis (see Everitt 1974) allows a researcher to determine the interrelationships between survey questions in much the same way as are used in a Factor Analysis. However instead of assuming that the relationship between the question responses is linear, we were able to use a measure of similarity between the responses that does not assume any ordering or cardinal relationship between the categories for the responses.

Using a hierarchical clustering routine and a complete linkage criterion for combining clusters, we clustered the questions into groups as described by the dendrogram in Figure 4.6. From the cluster analysis we find that N14 (*computer based teaching*) and N11 (*helpful feedback*) are placed in their own clusters when most of the other questions have been placed in groups. Also that N8 (*intellectually stimulating*) and N10 (*motivated to learn*) form a cluster that remains distinct from most other questions until the formation of four clusters.

From our analysis we can conclude that the CEQ and the TQI will measure different things, and that this relationship can be measured in a number of ways. The Item-Response Analysis demonstrates that the basic questions on the CEQ for teaching quality result in different scores from the ones found for the New TQI formed from the composite of the TQIs used in Australia. However we do find that N10 (*lecture motivation*) and N11 (*providing feedback*) are TQI questions that are closest to the CEQ Good Teaching Scale questions (C1-C6).

The Factor Analysis showed that most of the variation in the responses for the CEQ can be quite tightly grouped, so that one might conclude that they measure a different type of response from the TQI. However we can identify some exceptions. C13 (*the overall quality question*) appeared to have a stronger relationship with the TQI when measured by the first component dimension. But only C5 (*lecturers worked hard to make lectures interesting*) and C4 (*lectures were good at explanations*) were found to be closely related to the TQI.





**Figure 4.6:** Dendrogram using Complete Linkage and the Chi-square test for the equality of two sets of frequencies as the measure of distance.5.

The Cluster Analysis shows the most striking differences between the two types of surveys. In this analysis, we find that *N11* (providing feedback) and *N14* (use of computer materials) are the most dissimilar to the other questions as was also found when we examined the fit-statistics from the item person map.

#### 4.4.4 Regression Analysis

We also performed a regression analysis of these data on the Good Teaching Scale (GTS) which is the average of the responses to questions *C1* to *C6*. In a regression of the responses to all the other questions in the new TQI and including fixed effects to control for the subject and the institution, we were able to estimate the following parameter values.

In this case, we were interested in whether students’ responses to questions about the subject they were taking influenced their responses on a set of questions on their perceptions of quality of teaching for their whole course of study. The other variables left out were the constant and dummies for which subject or institution they attended.

From Table 4.6 we note that the most closely related questions to the CEQ Good Teaching Scale response were those that related to the lecturer motivation (*N10*), feedback (*N11*) and “good at explaining” (*N2*). With questions relating to organization and subject matter knowledge determined to be as less informative. However, since many of these questions request similar information, a question coefficient may not be significant because another question addresses the same or a related topic. From both the factor analysis and the cluster analysis we found that the feedback question (*N11*) appears to elicit a different type of response from the other questions in this set and we can conclude that this is an important factor in course evaluation. The motivation question (*N10*) has been closely aligned with the intellectual stimulation question (*N8*) which was insignificant in the regression



analysis, probably due to the possible confusion with the terminology of the question. It was also the case that (N2) good explanations and (N9) gaining a good understanding were closely related, however, although N2 was a significant factor, N9 was listed as insignificant.

Survey Questions impact of the CEQ GTS measure		Coeff	Sig
N10	The lecturer motivated me to learn	0.794	***
N11	I received helpful feedback on how I was going in this subject	0.740	***
N2	The lecturer was good at explaining the subject matter	0.632	***
N13	Lecturers showed an interest in the academic needs of the students	0.475	**
N15	The lecturer presented an appropriate amount of material for the time available	0.398	**
N4	The lecturer communicated enthusiasm for the subject area	0.342	**
N6	The lecturer has been available to discuss problems and questions	0.281	*
N5	The lecturer treated students with respect	0.248	
N8	This subject was intellectually stimulating	0.195	
N12	The lecturer clearly explained what I was required to do in assessment items	0.127	
N1	I had a clear idea of what was expected of me in this subject	0.110	
N9	The lecturer assisted me in gaining a good understanding of the subject matter	0.097	
N3	The lecturer was well organised	0.095	
N14	There was effective use of computer-based teaching materials in this subject	-0.075	
N7	The lecturer knew the subject matter well	-0.253	
* p<.05, ** p<.01, *** p<.001, robust standard errors			
The institution and subject fixed effects have not been reported here.			
R <sup>2</sup> = .58, 923 observations (only observations with complete data are used).			

**Table 4.6:** The regression parameter estimates of the responses to the new TQI on the CEQ questions.

## 4.5. The analysis of the institution specific questionnaires.

### 4.5.1 Introduction

This analysis explores relationships between institution-specific teaching quality instruments (TQIs) and the nationally administered Course Experience Questionnaire (CEQ). The analysis was replicated across four universities, each with a different TQI.

In broad terms, TQIs are designed for continuous improvement whilst the CEQ is used for summative external evaluation. This being the case, it is imperative that particular TQIs are empirically aligned with CEQ. If institutions undertake internal improvement activities in response to their internal TQI and these TQI are unaligned with the metrics used for external quality monitoring, then the internal or external systems or their linkages are corrupted. Of course, institutions may look to enhance aspects of education that are not considered by external quality monitoring activities. But on those measures where there are overlaps, one would expect a relatively high-level of relationship. Clearly, it would be counterproductive if institutions enhanced the quality of their teaching and learning and were penalised for this. Equally, it would be perverse if institutions that did not enhance the quality of their teaching, but still received recognition for so doing.

Tables 4.3 to 4.4 above present the items from the CEQ considered in this analysis. Table 4.7 below presents items for each of the four institutions along with their



labels.<sup>11</sup> The response categories provided on the survey forms are not shown in these tables, but they are worth noting in order to register their diversity. Flinders University students are asked to register their responses using a seven-point rating scale of strongly disagree, disagree, mildly disagree, undecided, mildly agree, agree and strongly agree. University of Melbourne students respond using the following scale: “strongly disagree”, “disagree”, “neither agree nor disagree”, “agree” and “strongly agree”. The University of Tasmania questionnaire uses the same form as The University of Melbourne except that the middle category is labelled ‘neutral’. The eight-point University of Wollongong rating scale is most extensive: strongly disagree, disagree, mildly disagree, mildly agree, agree, strongly agree, unable to judge and N/A. In this analysis, item-level analyses are conducted using the original metric whilst scale analyses are based on a recoding of all scores to a value ranging from 1 to 5.

#### *4.5.2 Investigating construct invariance in institutional specific TQIs*

Before comparing scores from the sets of items, it is important to determine whether the items themselves measure the same phenomena. If institutions’ own instruments measure things different from those on the CEQ, then driving internal change in ways that register as external improvement could be difficult. The key question is, therefore, whether the institutional surveys and the CEQ items measure the same constructs.

A series of covariance analyses were conducted to test this proposition. The first looked at whether the CEQ and institution-specific items displayed high levels of consistency when scaled together. The second looked at whether the two groups of items loaded on a single underpinning factor. Finally, confirmatory psychometric modelling was undertaken using Rasch item response modelling to test whether the items provided uni-dimensional measurement of a single construct. This was assessed by reviewing how the items distribute along a single variable, the relative difficulty of each item, and by reviewing mean square statistics which expose the fit of the item to the variable.

In addition to the analysis of construct invariance, a series of analyses were conducted to test the relationship among the CEQ and TQI scale means. Whilst progress to analysis of the empirical relationship between national and institutional mean scores if the instruments are not measuring the same construct may be questioned, this relationship is frequently explored in practice and it is worth doing so here.

---

<sup>11</sup> Figures C.1 to C.4 in Appendix C of the Resource Document for this report provide the questionnaires used for each university and they include the form of the local questionnaire as well as the new TQI and the CEQ items.



### ***Flinders University***

F1: Activities within the subject provided relevant learning experiences  
F2: I understood the concepts presented in this subject  
F3: The subject content was presented at an appropriate pace  
F4: The subject content was presented at an appropriate level of difficulty  
F5: The teaching materials and resources were helpful in directing my learning  
F6: Teaching materials and resources were culturally inclusive  
F7: This subject helped me develop my thinking skills (e.g. problem solving, analysis)  
F8: My ability to work independently has increased  
F9: I understood the assessment requirements of the subject  
F10: I received useful feedback on my learning  
F11: I was able to access quality support (e.g. from lecturers, other students, the university) when appropriate  
F12: Overall I was satisfied with the quality of this subject

### ***The University of Melbourne***

M1: I had a clear idea of what was expected of me in this subject.  
M2: This subject was well taught.  
M3: This subject was intellectually stimulating.  
M4: I received helpful feedback on how I was going in this subject  
M5: In this subject, lecturers showed an interest in the academic needs of the students.  
M6: I felt part of a group of students and lecturers committed to learning in this subject.  
M7: There was effective use of computer-based teaching materials in this subject.  
M8: Web-based materials for this subject were helpful.  
M9: Overall, I was satisfied with the quality of the learning experience in this subject.

### ***University of Tasmania***

T1: The unit addressed the learning outcomes stated in the Unit Outline  
T2: The criteria for each assessment component were clearly identified  
T3: The workload in this unit was appropriate  
T4: There was reasonable opportunity for interaction with lecturers  
T5: I was given useful feedback on my assessment work  
T6: Submitted work was returned to me in a reasonable time frame  
T7: The unit stimulated my interest in the subject area  
T8: I gained a good understanding of the subject matter  
T9: I enhanced my skills in this unit  
T10: The unit was well taught

### ***University of Wollongong***

W1: In this subject the learning objectives were made clear to me  
W2: The assessment criteria were clearly stated at the beginning of the subject  
W3: Feedback on my work was provided to me in time to prepare for other assessment tasks  
W4: This subject helped me gain a better understanding of an area of study  
W5: My learning in this subject was well supported by access to lecturers  
W6: My learning in this subject was well supported by access to other assistance  
W7: My learning in this subject was well supported by learning tasks  
W8: My learning in this subject was well supported by learning resources  
W9: My learning in this subject was well supported by eLearning (if used)  
W10: Overall I was satisfied with the quality of this subject

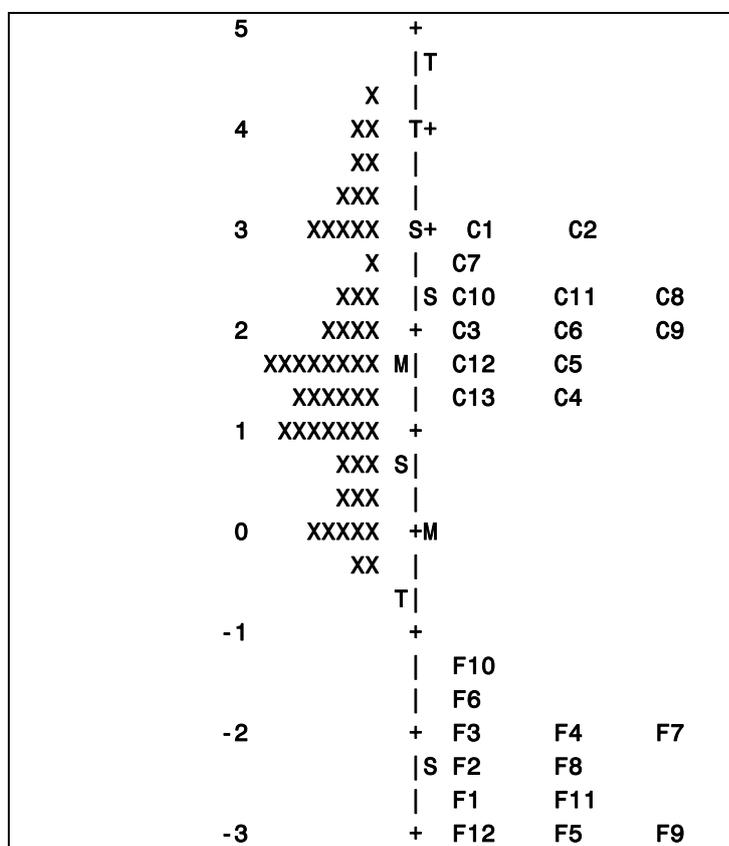
**Table 4.7:** Institution-specific items



In line with the design of the study, each analysis was replicated separately for each institution. Whilst some cross validation is afforded via this replication across institutions, it should be noted that the number of respondents per institution was relatively low. The total number of responses was 1022, with these being distributed there were 55 for Flinders University, 590 for The University of Melbourne, 211 for the University of Tasmania, and 166 for the University of Wollongong. Along with the small sample sizes, it is important to note that for this study data were only collected from students in the business field of education.

### 4.5.3 Flinders University

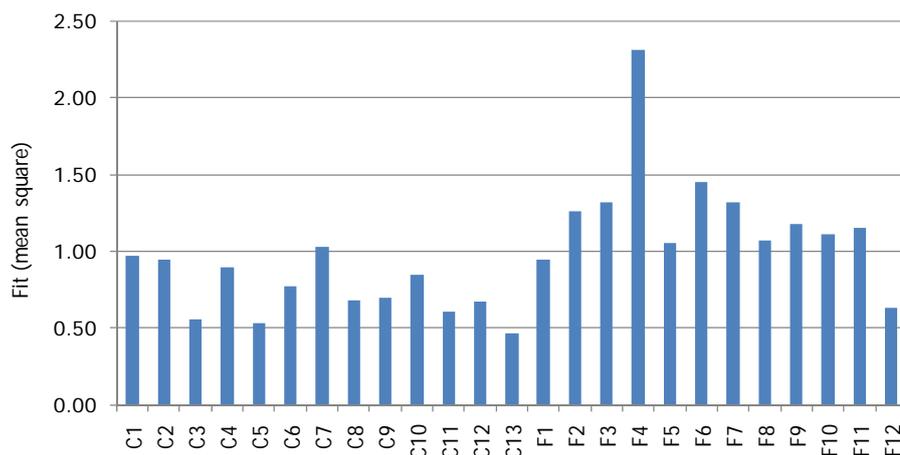
Figure 4.7 presents the variable map produced from Rasch analysis of the CEQ and Flinders University items. It shows the distribution of the items on the right-hand side of the variable, and the distribution of students' responses on the left-hand side as in Figure 4.2.



**Figure 4.7:** Item person map: CEQ and Flinders University items

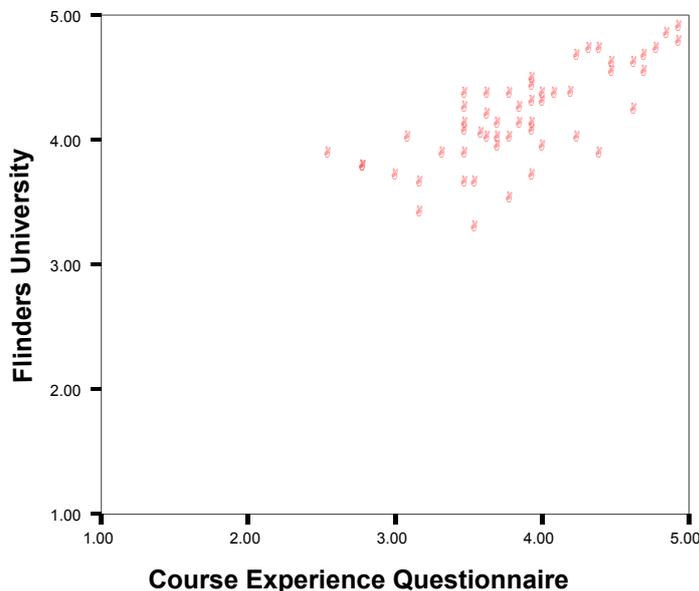
The item distribution in this case indicates that it is bi-modal, thus we can conclude that the CEQ and Flinders University instruments are largely measuring different constructs. The institution-specific items are considerably closer to each other than those of the CEQ, and do not appear to scale in an integrated fashion along the common latent variable as we found in the case of the new TQI as shown in Figure 4.2.





**Figure 4.8:** Fit (mean square): CEQ and Flinders University items

Figure 4.8 shows the mean square fit statistics arising from the Rasch analysis. These have an expected value of 1.0, and figures greater than 1.3 are conventionally read as reflecting a random relationship between the item and latent variable. The CEQ items show good fit to the variable. Flinders University items, by contrast, show a loose connection with the variable, particularly item F4, ‘The subject content was presented at an appropriate level of difficulty’ was found to have a very high value for the “fit” measure.



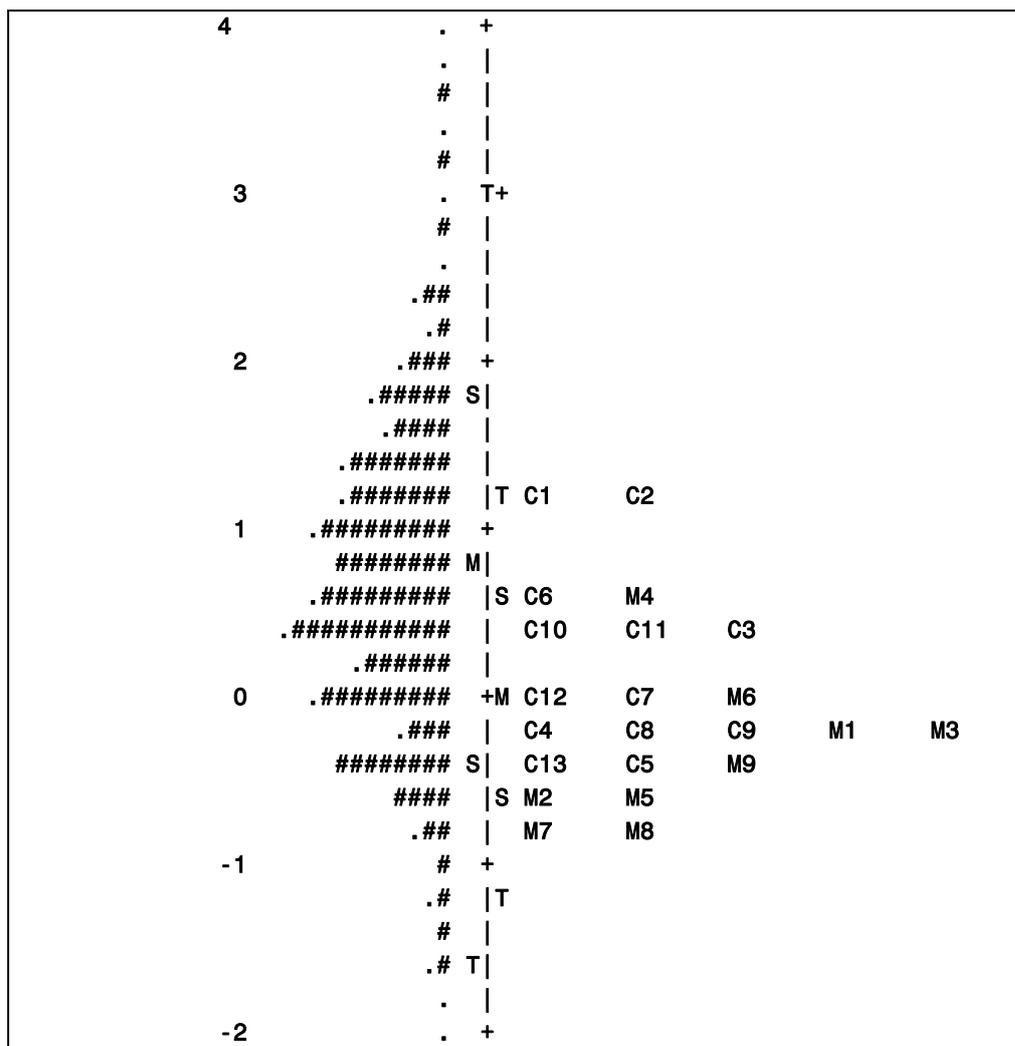
**Figure 4.9:** Scale mean score relationships: CEQ and Flinders University

Figure 4.9 charts the relationship between the mean scale scores for the student responses from the CEQ and the average of the Flinders items. The correlation between these was 0.72 and we did find that all the averages were in the positive response category. The plot confirms a relatively loose relationship between the external and internal metrics, particularly at the lower end of the distribution.



### 4.5.4 The University of Melbourne

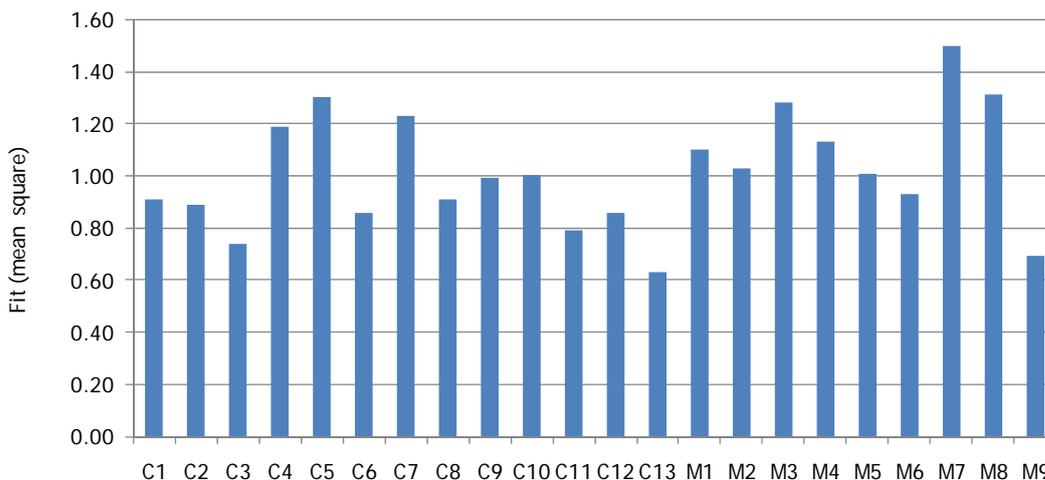
The presentations given above are repeated here for The University of Melbourne. Figure 4.10 shows that most of the national and institution-specific items do work together to map a common underpinning variable much as we found in Figure 4.2 where we assessed the relationship between the CEQ and the new TQI.



**Figure 4.10:** Item person map: CEQ and The University of Melbourne items

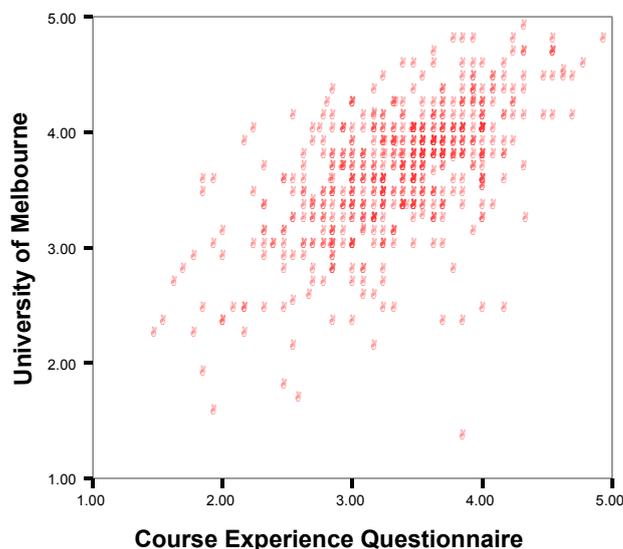
Figure 4.11 shows those items that contribute strongly to this common variable – exceptions include C5, M3, M7 and M8.





**Figure 4.11:** Fit (mean square): CEQ and The University of Melbourne items

Figure 4.12 confirms that the two instruments are closely related. Whilst there are cases in which students provided a higher global than subject-specific rating, the overall relationship is relatively linear. This is not strongly supported by the linear correlation between the measures, however, which is 0.62. In particular, there appear to be some outlier responses where the averages from The University of Melbourne survey questions are much lower than the CEQ responses. In particular, we observe cases where low average responses to the subject are not reflected in low responses to the CEQ. Thus indicating that these responses are more directed to the course as a whole and these respondents were ignoring possible difficulties they perceive with the subject they were taking.



**Figure 4.12:** Scale mean score relationships: CEQ and The University of Melbourne



### 4.5.5 University of Tasmania

The relationship between the CEQ and University of Tasmania TQI appears relatively robust (see Figures 4.13 and 4.14). The items which do not fit a single common variable include C7, T3 and T6.

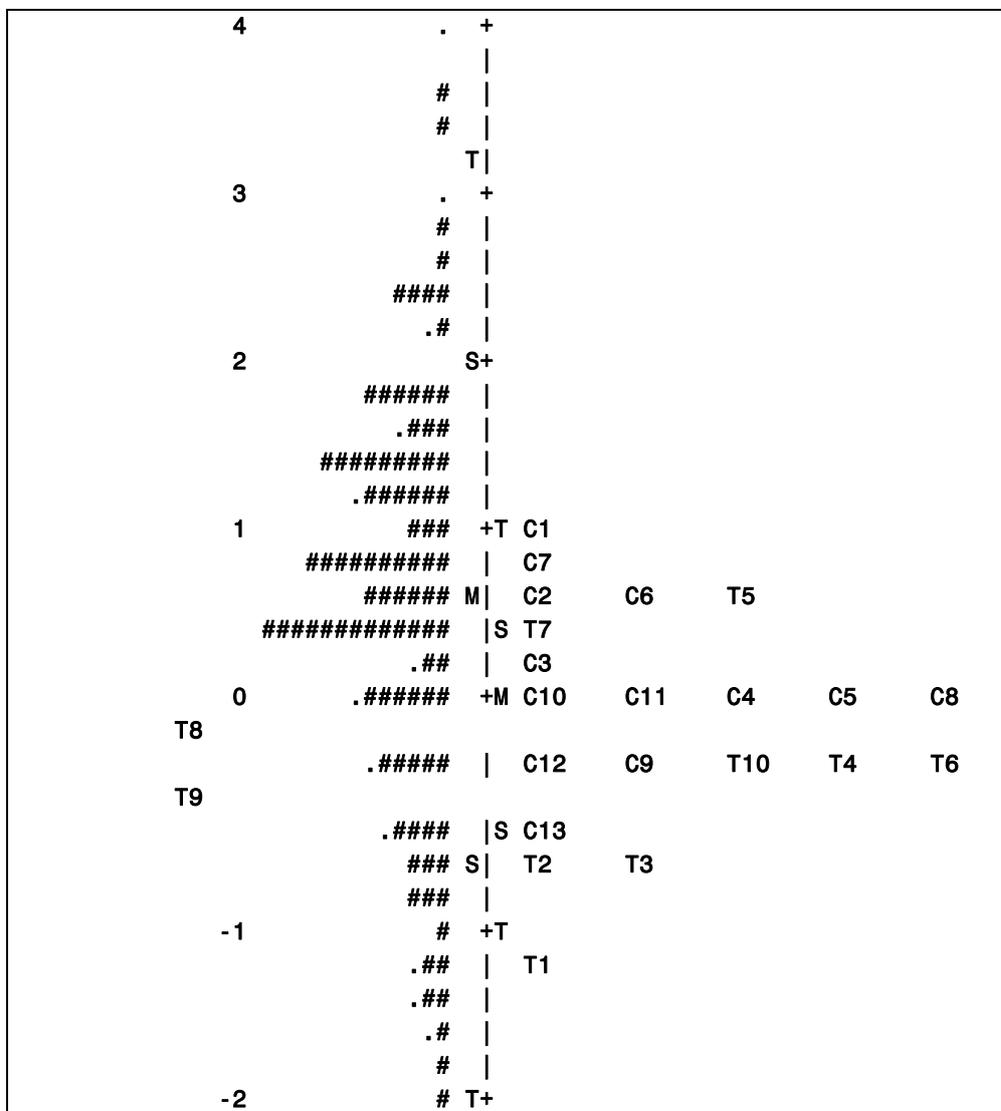
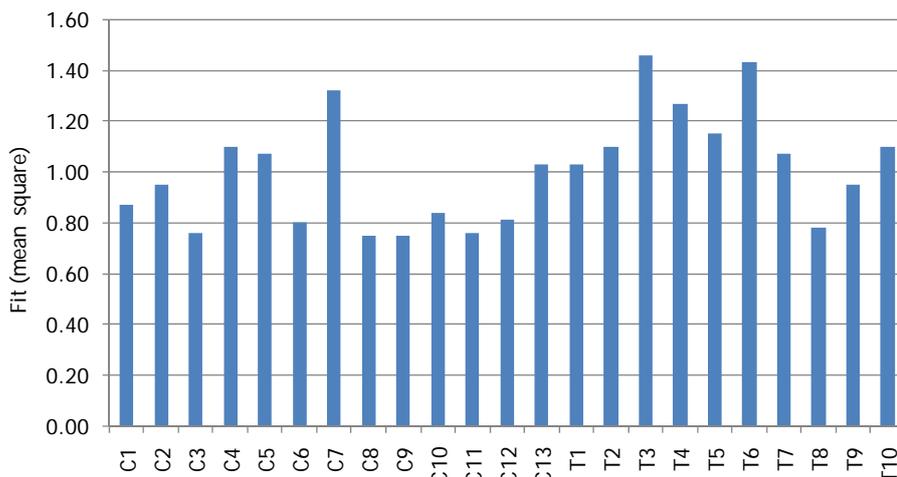


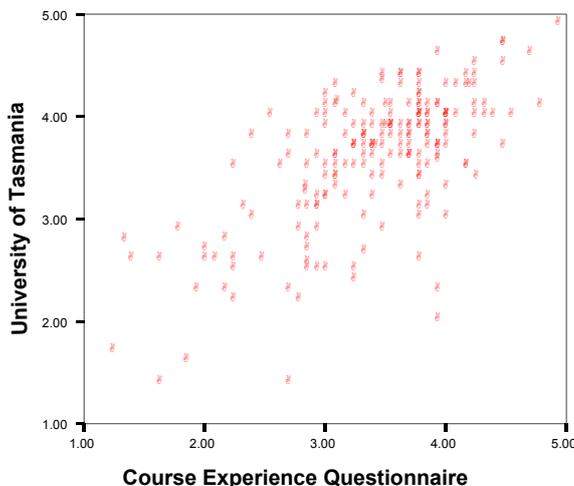
Figure 4.13: Variable map: CEQ and University of Tasmania items





**Figure 4.14:** Fit (mean square): CEQ and University of Tasmania items

In addition, there appears to be a rough but relatively linear relationship between the CEQ and University of Tasmania TQI scale means (Figure 4.15). The correlation between these scores was 0.69. Again from this scatter plot we note, as was the case of other institution specific TQIs, that some low local average measures are not matched by the CEQ where as the high ones are.



**Figure 4.15** Scale mean score relationships: CEQ and University of Tasmania

#### 4.5.6 University of Wollongong

Items in the University of Wollongong TQI appear to have the lowest relationship to the CEQ. As with Flinders University, the variable map shows that these items cluster together rather than spreading out to measure a common variable (Figure 4.16).

From Figure 4.17 it can be seen that almost all of the University of Wollongong items have mean square fit statistics that lie beyond the acceptable limit of 1.3 – the exceptions are W6, W7 W8 and W10.



The correlation between the average CEQ and average institutional scores is 0.62. Figure 4.18 shows that the relationship for the 166 scores under analysis looks quite diffuse with a number of outliers that would indicate that a number of students perceived that the items in each of these questionnaires related to different aspects of the subject and course they were taking.

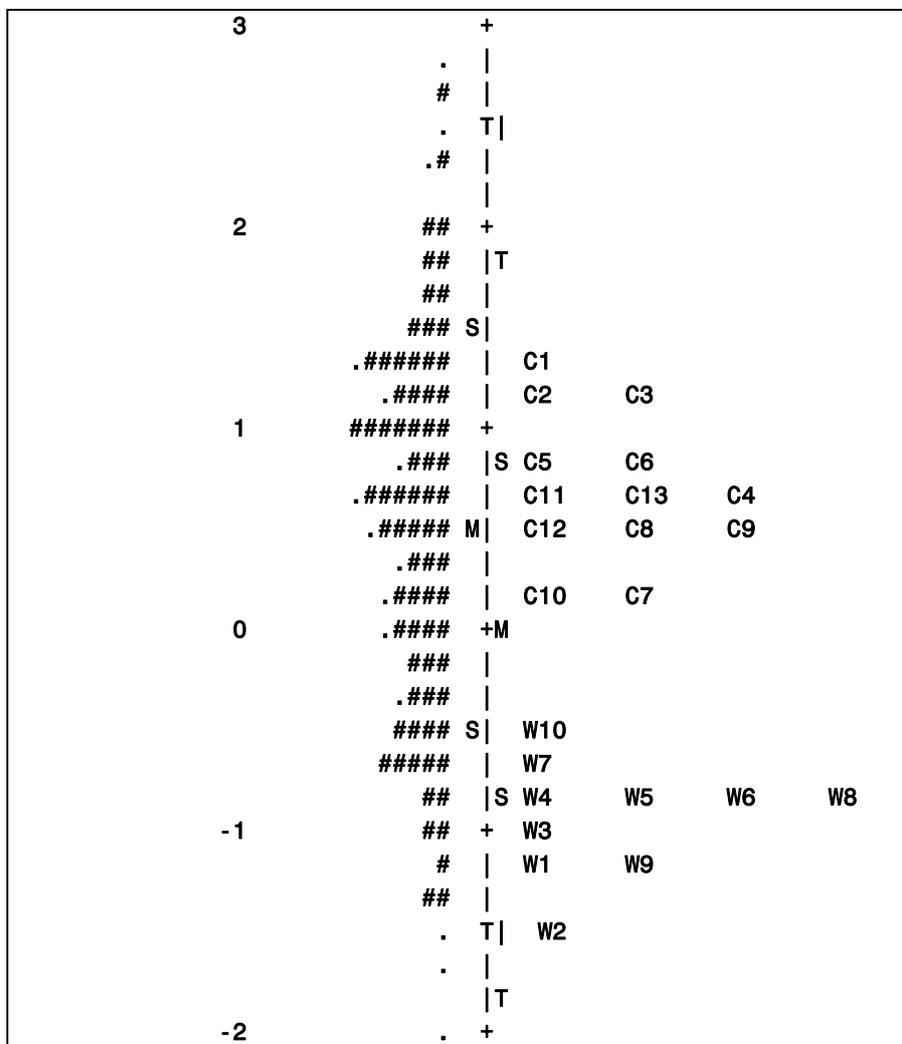


Figure 4.16: Variable map: CEQ and University of Wollongong items



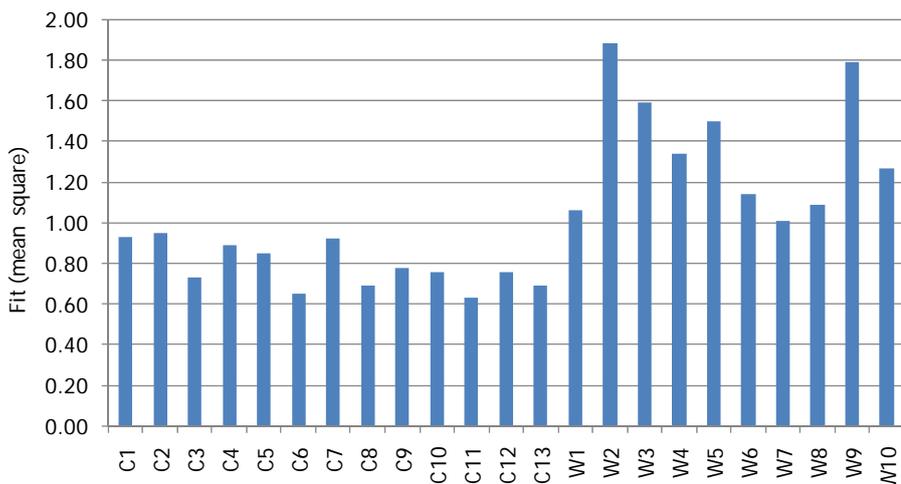


Figure 4.17: Fit (mean square): CEQ and University of Wollongong items

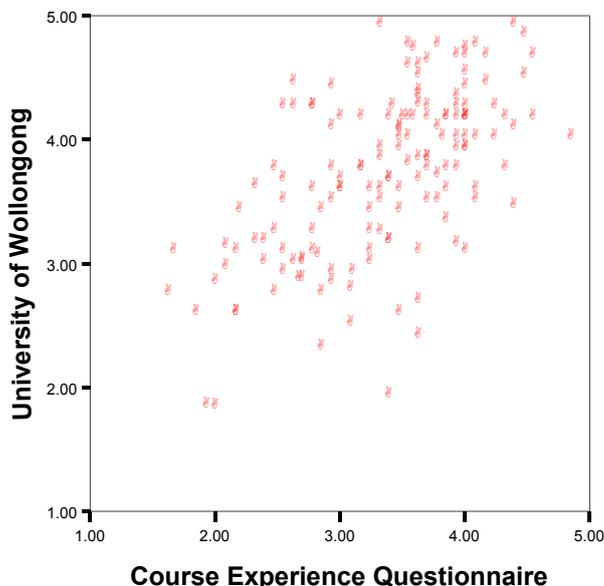
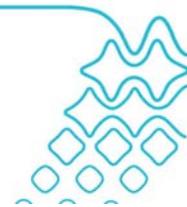


Figure 4.18: Scale mean score relationships: CEQ and University of Wollongong

### 4.6 Discussion

In this chapter we have demonstrated how one can measure the degree to which generic TQIs may indicate how a response on the CEQ can be formed by investigating the relationship between the responses to both types of questions. However, we also have shown that this mapping may be somewhat incomplete. Overall, these results suggest that there is a relatively low degree of association between the CEQ and TQI. Whilst based on a relatively small amount of evidence, the broad observation holds across various contexts and instruments.

There may be several reasons why the CEQ and the TQI items appear to measure different constructs. For example, the frame of reference for the TQIs may be



different from the CEQ's. The TQIs seek students' perspectives on a single subject, whereas the CEQ requests an aggregate perspective on the overall course. Furthermore, a student's current course experience may not translate to their impressions three to six months after they are finished the course when they complete the CEQ. However, here we have asked 1<sup>st</sup> and 2<sup>nd</sup> year students what they think of a course that may have one or two more years to complete. In principle, if the instruments are measuring the same constructs, the measurement process should not be confounded by this difference in the level of analysis.

Also in many cases the response scales used by the TQI and the CEQ did not align. Variations include the number and labeling of response categories, the presence and positioning of 'not applicable' categories, and the wording of the instrument. The wording is likely to introduce the most noise into the response process, for respondents who complete many such forms may be unlikely to check the precise labeling of the categories or their understanding of items before responding. Although all categories were labelled in a uniform way for the current analysis, this alone may have given rise to the response interference effect in question.

Most pointedly, it may be that the instruments do indeed measure different constructs. This is likely given that the TQI instruments have been developed at different times for different purposes, and because many institution-specific forms have not been psychometrically validated.

In sum, we find that the CEQ and the TQI each measure different things and that anticipating how student responses will change from subject or class responses to course or degree responses may involve factors that are not measured by the end-of-subject surveys.

In Chapter 5 of this report, we investigate responses by individuals to the CEQ. We demonstrate that other factors may be much more influential than those measured by the TQIs. These include: finding employment, choosing further study, grades achieved, expectations of a subject, and other demographic factors, and which seem to be instrumental in their final assessments.



## Chapter 5. The analysis of the response to the Course Experience Questionnaire (CEQ) as determined by the history of subjects taken by the respondents.<sup>12</sup>

### 5.1 Introduction

This chapter presents findings from a study of the influences on a set of CEQ response scores from students who completed courses at The University of Melbourne. Of the four institutions considered in this project, Melbourne was the only institution in which we were able to match the CEQ responses to the experiences of the students who completed the survey and thus attempt to determine the degree to which the CEQ responses are influenced by the both the characteristics of the student and their experiences whilst taking subjects in the University. In this case we had access to a set of subject specific TQI results over most of the period of their course of study and their responses to the CEQ. We were therefore able to establish the degree to which these results may predict the subsequent CEQ response. In addition, this analysis was extended to faculties other than Business and Economics to allow for more variation in the responses across differing instructional styles.

In this chapter, we first describe the assembly of data used in the analysis along with an analysis of the distribution of the responses to the CEQ by course of study. We accompany this with an analysis of the data used here to demonstrate that the seven aggregate scales in the CEQ appear to measure different characteristics.

Secondly, we describe the TQI used at The University of Melbourne during the time of the data collection and demonstrate the presence of department and Faculty specific TQI response patterns. We also include an analysis of potential response bias by the examination of how subject characteristics might predict the propensity for students to respond to the CEQ.

The third section of this chapter describes the mixed regression models used to analyse the individual responses to the CEQ with special attention to the responses to the Good Teaching Scale. In order to perform this analysis, we needed to establish those factors in the students' experiences that might have influenced their responses. We subdivided these influences into those that described the subjects taken by the student, such as the average TQIs in their subjects, and those factors that are individual to the student such as their rank in the subjects they took. Also included in this section is a description of the seven models we used to account for systematic influences on the students' whole course of study experiences as determined by the subjects they completed. We present the results here with particular emphasis on the impact of each subject TQI on the subject specific models of TQI described in Section 5.2.2.

---

<sup>12</sup> Early versions of this chapter have been presented at Department of Economics, Southern Methodist University, Dallas, TX, January 8, 2009. The 7<sup>th</sup> Hawaii International Conference on Statistics, Mathematics, & Related Fields 2009, January 15, 2009, and The Forum on Quantitative Analysis Of Teaching And Learning In Higher Education In Business, Economics And Commerce, The University of Melbourne, February 6, 2009.



Finally, we examine how the model used in Section 5.3 can describe the responses to the other CEQ scales. We examine the responses to these factors to the first model specification after showing how these models fit each CEQ score.

## 5.2 Data

The relationship between the various data sources used in this analysis is described in Figure 5.1. The majority of the information has come from three sources: The CEQ and GDS (Graduate Destination Survey) which are filled out at the same time;<sup>13</sup> the records of students who filled out the surveys with all identifying information removed except marks and subjects; and the average values for the questions asked in the Quality of Teaching Survey (QOT) administered at The University of Melbourne.<sup>14</sup> The chart below indicates how these data sets are related and the process by which they are combined to create a student record level data set with which we can perform the analysis.

We first constructed a history of subjects taken by each student who completed the CEQ/GDS including their mark in each subject. This was then matched to the subject specific QOT results which were available as average responses by question. These averages were then matched to each student's history to indicate the quality of the subject they had taken. In addition, to the QOT averages, we also used the set of all student enrolments to establish the distribution of the marks in all subjects taught during this period. This file was then matched to the student records of those who had responded to the CEQ/GDS in order to compute the rank of the students in each subject they took. In all we had 14,728 course surveys

### 5.2.1 CEQ/GDS data

The first data set includes responses on the CEQ/GDS for the four years 2002-2005.<sup>15</sup> These responses are for all students who returned forms at The University of Melbourne. Note that 10,433 students reported on their experiences in 14,728 separate courses. Some students had taken more than one course, therefore Table 5.1 lists the frequency of the number of courses by student.

Number of courses	Number of Students	%
1	6594	44.77
2	7066	47.98
3	588	3.99
4	480	3.26

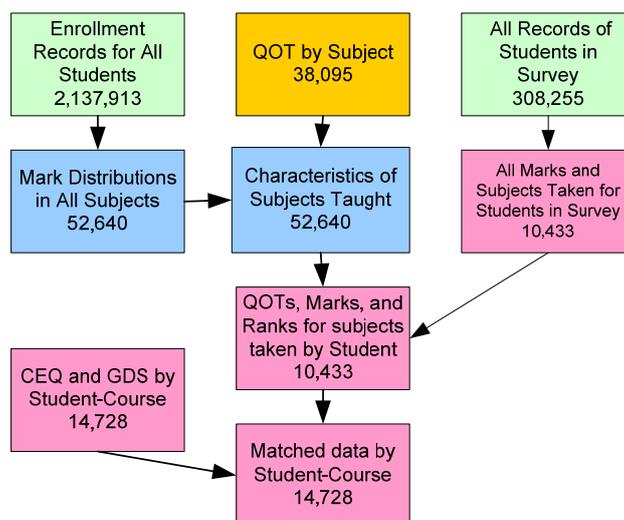
**Table 5.1:** The distribution of the number of courses per reporting student.

<sup>13</sup> The coverage of the GDS is a bit wider than the coverage of the CEQ. This is due to the use of telephone surveys for the CEQ versus the GDS where phone contact was not needed.

<sup>14</sup> Note that the use of the acronym QOT is unique to The University of Melbourne and that many other acronyms are used both in Australia and abroad. In fact CEQ is used in the US to denote the subject specific surveys because the term "course" is used for what is referred to here as subject. A list of the acronyms used in Australian tertiary institutions is given in the Appendix A in the Resource Document for this report.

<sup>15</sup> A description of the survey items used from the CEQ/GDS is given in Appendix D in the Resource Document for this report.





**Figure 5.1:** The interrelationship of the data used to perform the analysis.

These courses were taken in a number of different faculties. The distribution of courses taken by the students who completed the CEQ/GDS survey by faculty and year are listed in Table 5.2 listed below.

<b>Faculty</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>
<b>Agriculture</b>	155	134	107	163
<b>Architecture</b>	161	148	150	176
<b>Arts</b>	680	734	840	874
<b>Economics &amp; Commerce</b>	384	341	441	524
<b>Education</b>	199	179	154	182
<b>Engineering</b>	284	238	265	373
<b>Law</b>	185	154	179	217
<b>Medicine</b>	198	189	232	325
<b>Music</b>	59	66	82	91
<b>Science</b>	1036	929	1137	1337
<b>Veterinary</b>	35	33	38	39
<b>Vic College of Arts</b>	84	41	78	78

**Table 5.2:** The distribution of courses by faculty and year taken by reporting student.

The CEQ response data were categorized using a 5 point Likert scale positive responses (most positive = 5 and least positive = 1). In order to establish how the responses to these questions compared with each other, we conducted a cluster analysis in which the distance metric is defined as the chi-square statistic for the test that the distributions of the responses to the questions are the same.<sup>16</sup> The dendrogram in Figure 5.2 shows the hierarchical clustering of the questions using the complete linkage method in which the distances between clusters are defined by the maximum of the distances among all the possible comparisons in each cluster.

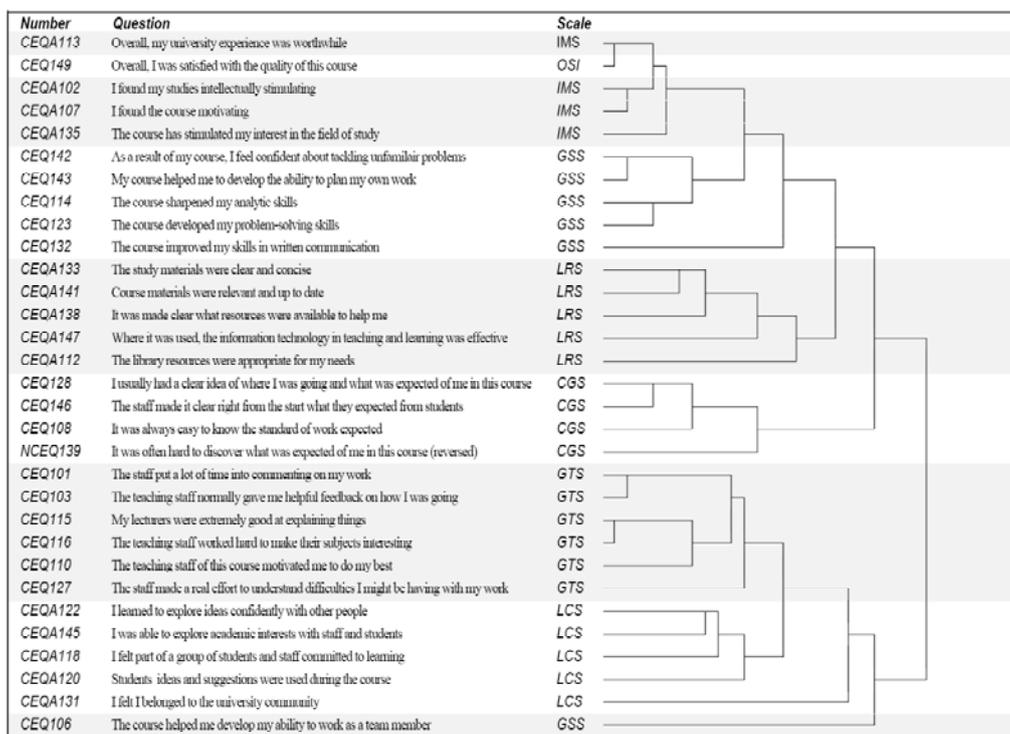
<sup>16</sup> The set of distances computed for this analysis are listed in the Appendix E in the Resource Document for this report..



Also included in this dendrogram is the designation as to whether the particular question is included in a particular CEQ scale. These are defined in Table 5.3 shown below. Note that the clusters defined coincide with the scale definitions for all but one question in the Generic Skills scale (CEQ106 - The course helped me develop my ability to work as a team member). This indicates that using the scale scores (defined as the average of the responses to the questions included in the scale) does not lose much information that may be found in the separate questions. Thus we can concentrate on these scales and not concern our analysis with the individual question responses.

<b>Mnemonic</b>	<b>Name of CEQ Scale</b>
<b>GSS</b>	Generic Skills
<b>CGS</b>	Clear Goals and Standards
<b>GTS</b>	Good Teaching
<b>IMS</b>	Intellectual Motivation
<b>LCS</b>	Learning Community
<b>LRS</b>	Learning Resources
<b>OSI</b>	Overall Satisfaction

**Table 5.3:** The classification of CEQ questions by scale.



**Figure 5.2:** The Dendrogram of the responses to the CEQ as reported by students at The University of Melbourne



### 5.2.2 The Quality of Teaching (QOT) survey data

At The University of Melbourne teaching surveys have evolved since their inception in 1995, however some questions have remained largely unchanged. Since we required a series of teaching quality standards that remained comparable over time, only those that remained constant were suitable for our use.

Mnemonic	Period of use	Question
CQ1	S1 1997 - current	I had a clear idea of what was expected of me in this subject
CQ1A	1994 S1 and S2	The objectives of the subject were made clear
CQ1B	1995 and 1996	The aims of this subject were made clear
CQ2	1994 - current	This subject was well taught
CQ2B	1994	The program assisted me to achieve the subject objectives
CQ3	1995 - current	This subject was intellectually stimulating
CQ3B	1994	The subject challenged my intellectual ability
CQ4	1997 - current	I received helpful feedback on how I was going in this subject
CQ4B	1994 Semester 2	I have received useful feedback on previous questionnaires completed in my course
CQ4C	1995-1996	The outcomes of previous questionnaires in this subject have been explained to me in a constructive manner
CQ5	1995 - current	In this subject, teaching staff showed an interest in the academic needs of students
CQ5B	1994	My ability to undertake independent study was improved by taking this subject
CQ5C	1994 Semester 2	In this subject, the teaching staff are responsive to and supportive of the needs of students
CQ6	1997 - S1 2003	The volume of work in this subject was appropriate
CQ6B	Semester 2 2003 - current	I felt part of a group of students and staff committed to learning in this subject
CQ7	Semester 2 2003 - current	There was effective use of computer-based teaching materials in this subject
CQ7B	1998-1999	The multimedia-based technology helped me to learn effectively
CQ8	Semester 2 2003 - current	Web-based materials for this subject were helpful
CQ8B	1998-1999	I regularly made use of the information and materials made available by the teaching staff on the Internet
CQ9	Semester 2 2003 - current	Overall, I was satisfied with the quality of the learning experience in this subject
DQ7	2000 - S1 2003	The teaching and learning program in this subject enabled me to learn effectively
DQ8	2000 - S1 2003	The technical/administrative quality of the subject delivery was good
DQ9	2000 - S1 2003	I was satisfied with the academic support offered in this subject
IQ7	2000 - S1 2003	The computer-based multimedia programs helped me to learn effectively
IQ8	2000 - S1 2003	My learning activities in this subject regularly made use of the web
IQ9	2000 - S1 2003	I found it useful to access information and subject materials through the subject website
CQ4D	1995-1996	This subject helped develop my learning skills.

**Table 5.4:** The questions used on The University of Melbourne QOT

From Table 5.4 we note that only CQ1, CQ2, CQ3, CQ4 and CQ5 were consistently used over most of this period. A principal components analysis of the correlation matrix of these questions over all the subjects for which these survey results were available found that over 76% of the variation is explained by the first component and that CQ2 has a correlation of .92 with this linear combination of the other scores. For this reason, for the analysis conducted here, we will concentrate on the results of question 2 (CQ2) “This subject was well taught”.



In an earlier paper (Davies et al. 2007), it was found that many factors influence the results of CQ2 of the QOT for the Economics Department at the University of Melbourne. Among these it was found that the size of the subject enrolment, the year level, the year in which it was administered, and the response rate (the proportion of the enrolled students who answered the survey had an influence on the QOT scores. These factors are used in a series of regressions that were fit using data for separate departments with the following model:

$$y_{ijt} = \theta_j + \lambda_t + \beta_1 x_{ijt} + \beta_2 x_{ijt}^2 + \gamma_1 z_{ijt} + \gamma_2 z_{ijt}^2 + \varepsilon_{ijt}$$

Where  $y_{ijt}$  is the QOT Question #2 average response observation  $i$ , subject  $j$ , taught in year  $t$ , with enrolment  $x_{ijt}$ , and response rate  $z_{ijt}$ . The purpose of this regression is to estimate an expected QOT score for each subject. In this way we may determine if the QOT scores observed are out of the ordinary or not. Note that almost 88% of the department models resulted in a rejection of the hypothesis that the subjects have the same average QOT score within the departments (only 13 out of the 106 departments for which we had sufficient data to estimate the regression) after accounting for the other factors in this model. We also found that, once other subject differences have been accounted for, approximately 40% of the departments have QOTs that have changed over the years from 1996 to 2005. In addition, we found that in a number of departments, once the year level and subject have been accounted for, almost one third are influenced by enrolment size and more than 22% by the response rate.

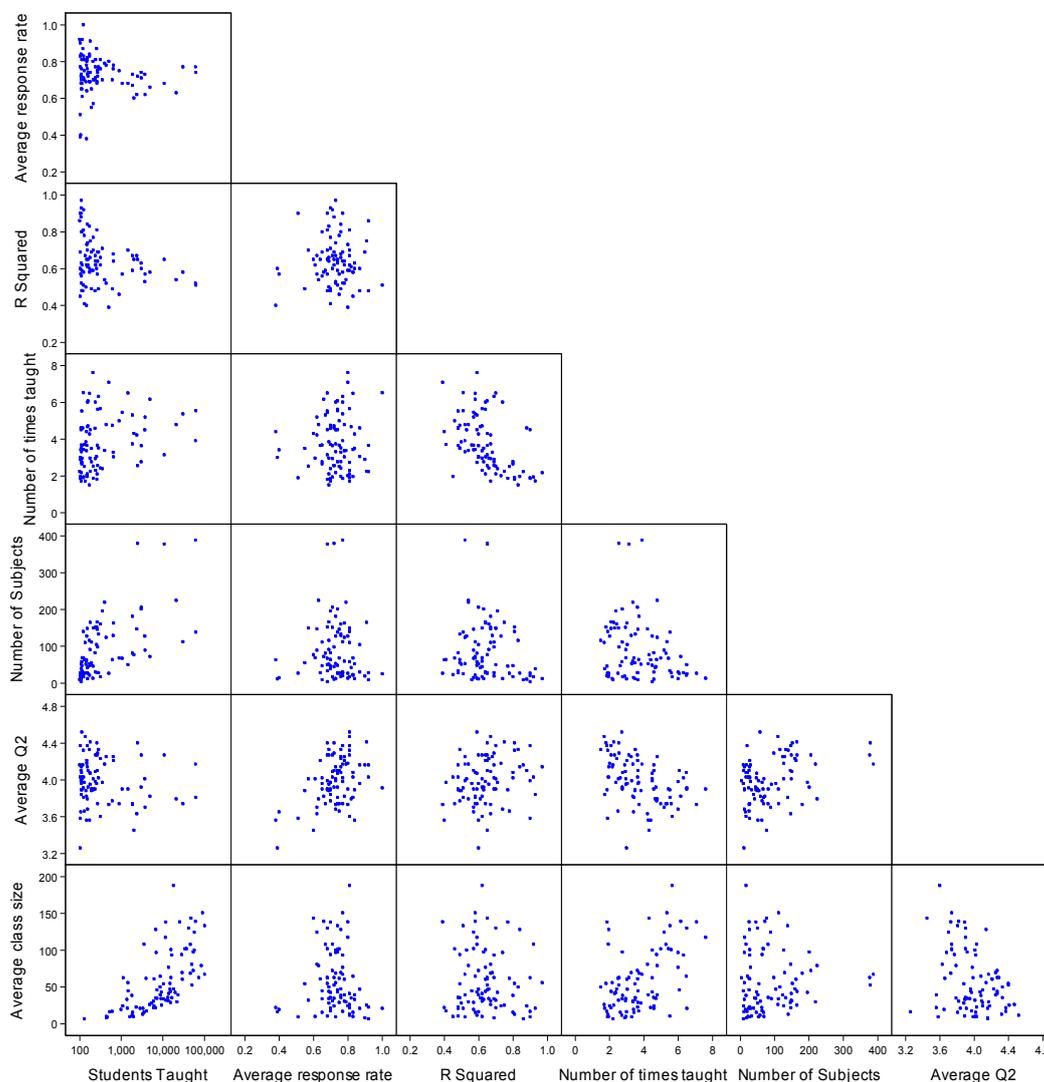
<b>Hypothesis to be tested</b>	<b>Probability of rejection % of Departments</b>	
$\beta_1 = 0$ parameter on enrolment	< .01	14.29
	.01-.05	26.67
	.05-.10	30.48
	> .10	100.00
$\beta_2 = 0$ parameter on enrolment squared	< .01	7.62
	.01-.05	18.10
	.05-.10	25.71
	> .10	100.00
$\theta_j = \theta_k, \forall j \neq k$ all subject parameters	< .01	79.25
	.01-.05	87.74
	.05-.10	88.68
	> .10	100.00
$\gamma_1 = 0$ parameter on response rate	< .01	7.62
	.01-.05	14.29
	.05-.10	22.86
	> .10	100.00
$\gamma_2 = 0$ parameter on response rate squared	< .01	7.62
	.01-.05	13.33
	.05-.10	19.05
	> .10	100.00
$\lambda_t = \lambda_s, \forall t \neq s$ all year parameters	< .01	11.32
	.01-.05	33.02
	.05-.10	40.57
	> .10	100.00

**Table 5.5:** The results of the tests of hypotheses of the parameters estimated in model (1).

Table 5.5 lists the results of F-tests for the influence of various parameters in the



department models. Figure 5.3 provides the scatter plot matrix of the summary statistics for each regression by department.



**Figure 5.3:** A set of scatter plots of the department averages.

From Figure 5.3 we can determine if there exist any bivariate interrelationships between the departmental average values. The  $R^2$  statistic indicates how well the model fits the data, the closer to one, the better the fit and more of the variation in the data explained by the model. Note, that the most prominent relationship with the average response to Question #2 is the response rate. From this plot, it appears that higher response rates were generated when the average responses for Question #2 were highest. In order to establish if we can find any partial relationships, we estimated a simple multivariate regression on these variables in order to condition these bivariate relationships. The results of a weighted regression are reported in Table 5.6. In this case, the weights are the inverse sample size to account for possible heteroskedasticity generated by the disparate number of subjects surveyed for each department.



<b>Variable</b>	<b>Coefficient</b>	<b>Prob.</b>
<b>Intercept</b>	1.554902	0.0000
<b>Average subject size</b>	-0.001475	0.0065
<b>Response rate</b>	6.701696	0.0000
<b>Response rate squared</b>	-4.114711	0.0000
<b>Number of times taught</b>	-0.041127	0.0027
R-squared	0.590104	Mean dependent var3.977614
Adjusted R-squared	0.570586	S.D. dependent var 2.242629
Included observations: 89		
Weighting series: inverse square root of number of subjects taught		
White Heteroskedasticity-Consistent Standard Errors & Covariance		

**Table 5.6:** A weighted regression over the average response to question #2 by department.

From Table 5.6 we note that different departments have differing average responses to the Question #2 that are determined negatively by their average enrolment size, positively (up to a point) by the rate of response, and negatively by the frequency with which the subjects have been taught. These conclusions can be reached from this table directly and indirectly for the response rate. Thus we can establish that gain from increasing response rates stops at rates of 0.75 or more, based on the confidence interval for the point at which the relationship between response rate and average Question #2 scores becomes insignificant. As with enrolment size, the impact of the frequency of teaching subjects is also negative – the more often a subject is taught appears to have a negative impact.

The implications of these results are that the average response to Question #2 can be modelled across departments as fundamentally different. This implies that students in different departments, and thus in different faculties, will expect different levels of teaching quality based on the characteristics of the subjects in those departments. That is, if a department has a large number of subjects that are taught over and over again with large enrolments, this model would predict that their average Question #2 scores would be lower. Also the response rate may well be low for these subjects given the tendency for low attendance in large compulsory subjects. These two factors predict that departments teaching “service subjects” in areas where the subject matter does not change due to professional certification restrictions, or the need to teach in course curricula that have a strongly hierarchical set of prerequisite subjects, will have lower responses on the QOT Question #2.

### 5.2.3 Who responds to the CEQ?

The enrolment data made available to this study and from all students at The University of Melbourne did not include an individual identifier that would allow the construction of a description of all the subjects they may have taken whilst they were at the university. Thus, although we were able to construct a student experience record for each student who filled out the questionnaire, we were unable to construct an equivalent student experience record for the students that did not complete the CEQ. In particular, we are unable to establish how the students who did not respond to the CEQs over the period of the data differed from those who did respond. If there were a systematic reason why only certain students responded, then we may have been able to conclude that the sample of CEQ responses does not reflect the true quality of the courses under evaluation since the sample observed is not representative of all students. This phenomenon is referred to as potential sample selection bias.



In order to gain some insight into the nature of the respondents to the CEQ and if there is evidence of potential sample bias, we performed an analysis by subject taught. This analysis does not establish the degree of selection bias, but does indicate if certain subject characteristics may result in a differential level of response. The analysis performed here employs a probit regression analysis in which the proportion of students in each subject who responded to the CEQ is listed as the dependent variable. To account for the fact that the subjects were measured from 1996 to 2005 and that the CEQ surveys were only available for the 2000 to 2006 period, we included variables to account for year and the level of the subject (i.e. 1<sup>st</sup> year – or post-graduate). In addition, we also used the distribution of marks and the results of the QOT surveys to measure the characteristics of the subjects. Theoretically, all students in every subject would have an equal likelihood of completing the CEQ surveys and thus, once we accounted for the year and level of the subject, the proportion of students in each subject should have been the same.

An important reason that they may not all have responded in the same manner may have been that they were unreachable when the CEQ is sent to them. This would be a problem for overseas students in particular. Thus one would expect that subjects that have a high proportion of overseas students would have a lower CEQ response rates than those with mostly domestic students. To account for this disparity, we also used a series of fixed variables to account for the department from which subject originated – based on the knowledge that some departments and faculties have a much higher proportion of overseas students than others. Therefore, in addition to fixed effects to account for the level of the subject and the year in which it was taught, we also included fixed effects for the department in which it was taught. The other variables of interest were the enrolment numbers, characteristics of the grade distribution, and the results of the QOTs administered in the subject. Table 5.7 below lists the results of the estimation:

<b>Variable</b>	<b>Estimate</b>	<b>Prob <math>\chi^2</math></b>
Intercept	-1.9856	0.0743
Proportion of marks close to 50 $\#(47-49)/\#(50-65)$	0.0750	0.0001
Proportion of marks close to 80 $\#79/\#(80-100)$	0.0360	0.0001
Range of marks	-0.0002	0.0055
Proportion of H1s	-0.0483	0.0001
QOT Q2 "This subject was well taught"	0.0152	0.0254
QOT Q3 "This subject was intellectually stimulating"	-0.0274	0.0001
QOT Q5 ".teaching staff showed an interest in the academic needs of students"	-0.0178	0.0109
Response rate on the QOT	-0.0023	0.8250
Enrolment /100	-0.2150	0.0001
Enrolment/100 squared	0.000069	0.0001
Times taught since 1996	-1.9856	0.0743
Semester*year	F-test	0.0000
Department	F-test	0.0000
Subject level	F-test	0.0000

**Table 5.7:** The results of a probit analysis to determine which factors influence the proportion of subject enrollees who ultimately respond to the CEQ. (prob  $\chi^2$  indicates the significance of the test that the parameter estimate is equal to zero)

From Table 5.7 we note that the presence of close marks in the range of passing marks (measured as the ratio of the number of marks between 47 and 49 to the



number of marks between 50 and 65), or at the top end the proportion of marks just below the H1 – high honours mark to the total of these marks (the ratio of the number of marks of 79 to the number of marks from 80 to 100). For both of these indicators we find that the parameters are significant and positive which indicates that the greater the proportion of close grades as defined by these measures, the greater the response rate by students in these subjects on the ultimate CEQ.

From Table 5.7 it can also be shown that some characteristics of the distribution of marks have a negative impact on the propensity for students that take these subjects to complete the CEQ. The range of marks (the difference between the maximum and minimum) has a negative impact on response rates where the greater the range, the lower the rate of response. The proportion of H1 marks in the subject also has a negative impact on the responses from this subject. The greater the proportion of high honours marks (80 or above), the lower the likelihood of the students in the subject completing a CEQ.

We then examined the impact of the average response in the subject to Question # of the QOT ("This subject was well taught). From Table 5.7 we see that the higher this value, the greater the probability that the average student will have completed a CEQ. However when we included other average responses from the QOT, we found that both Question #3 of the QOT ("This subject was intellectually stimulating") and Question #5 of the QOT ("...teaching staff and responsive...") have the opposite impact. This implies that the greater the perception that teaching staff pay attention to students, the lower the chance that the students will have completed a CEQ. We also found an inverse relationship between the response rates to the CEQ and the QOT. One explanation for this may be that students felt powerless in subjects with low response QOT rates so when provided the opportunity they decided to complete the CEQ. Alternatively, low response rates (all these forms are paper based and filled out in class usually during the penultimate week of lectures) may be an indication of low attendance rates due to dissatisfaction with the lectures or the lack of additional information gained from attendance.

The enrolment numbers in a particular subject also seems to have an impact on the response rate to the CEQ. Note that we have allowed enrolment to have a nonlinear impact by including a quadratic relationship. In this case low enrolment numbers showed a negative but diminishing impact on CEQ response and thus have a relationship that can be described as U-shaped. We also included the number of times the subject had been taught over the entire period (whilst some subjects may have been taught over 20 times, the average was approximately 5). We found that the more frequently a subject had been taught, the lower the probability that students will fill out a CEQ. Such subjects are more likely to be prerequisites and therefore carry less interest for the student.<sup>17</sup>

The lower portion of Table 5.7 reports results of F-tests for equality of the fixed effects that were estimated for department, year-semester and year-level. From these tests we can reject the hypothesis that these fixed effects are equal to each other.

We can conclude from this analysis that the hypothesis that all students in all

---

<sup>17</sup> It was found in a separate regression to explain the average response to QOT Question #3 ("This subject was intellectually stimulating") after accounting for the fixed effects considered above, the coefficient on the frequency a subject had been taught, was significantly negative.



subjects have an equal probability of completing the CEQ can be rejected. There are two possible reasons for this. First, the difference may be due to students not finishing a degree and thus never being in the position of filling out the surveys – although, if this were the case, one would assume that the higher proportions of H1s (marks of 80 or above) would have had a positive impact on the completion rate and thus on the proportion of students who fill out a CEQ. However, from Table 5.7 this is not the case, as the higher the proportion of H1s, the lower the probability that the students in that subject will fill out the CEQ.

Alternatively, we may conclude that filling out the CEQ may in some way be related to students’ experiences whilst undertaking their studies. We found that the proportion of close grades for all students in the subjects had a positive impact on the likelihood that the students would fill out the CEQ. These close grades or taking subjects that are characterized by having close grades appear to generate a higher response to the CEQ when these students are surveyed.

### 5.3. Analysis of the response to the Good Teaching Scale (GTS) of the CEQ

In this section we discuss how individual responses to the CEQ may be influenced by the experiences of the graduates who respond. As shown in section 5.2.3, there are reasons to believe that respondents are not randomly deciding to fill out the form, but that certain types of experiences seem to generate more responses than others. In this section we will determine to what extent these experiences may drive the responses to the CEQ.

As shown in section 5.2.1, there were a number of individual questions in the current CEQ and, of these, a number of composite responses. This analysis focuses on responses to the group of questions referred to as the GTS or “Good Teaching Score”, in that one would expect that this score would most be closely related to the characteristics of subjects.

#### 5.3.1 The Model Specification

The model used was linear regression with an error structure that accounts for heteroskedasticity generated by the response style of the individual. The model estimated was in the class of mixed models. The specification used to fit to the individual responses was of the form:

$$\underbrace{y_t}_{\substack{\text{CEQ} \\ \text{score} \\ \text{"GTS"}}} = \alpha + \underbrace{\sum_{i=1}^K x_{ti} \beta_i}_{\substack{\text{Student} \\ \text{Experience} \\ \text{(Individual \&} \\ \text{Subject)}}} + \underbrace{\sum_{j=1}^M d_{tj} \lambda_j}_{\substack{\text{Course} \\ \text{Description} \\ \text{"fixed effects"}}} + \underbrace{\sum_{l=1}^L z_{tl} \gamma_l + \varepsilon_t}_{\substack{\text{Random} \\ \text{Component}}}$$

Where  $y_t$  represents the student response for the CEQ scale of interest,  $x_t$  indicates characteristics of the student  $t$  and includes their academic performance as well as characteristics of the subjects they enrolled in, and  $d_j$  represents the fixed effects to account for the course taken – these can also be course level (graduate or undergraduate), faculty, the department in which the majority of the subjects were taken, and finally the ultimate occupations or destinations of the students. In order to model these “fixed effects”, we used seven alternatives for the definition of  $d$ . The random component of the model is made up of  $\varepsilon$  which is distributed with a mean



of zero identically and independently according to the normal distribution, and occurrences of  $\gamma$  which are distributed normally with mean zero and multiplied by  $z$  which defines how the random component varies. In this case we have used the extreme response index to model the variances.

Response style has been shown to influence variation in responses to individual questions on questionnaires when surveying different groups of individuals (see Van Herk, Poortinga and Verhallen 2004). By response style we mean the preference for some individuals to be more unequivocal than others in their responses to a typical Likert scale. Thus in the case of the 5 point scales used in the CEQ, they may respond with more 1's and 5's than with 2's, 3's and 4's. This tendency can be determined for an individual by using their responses to all the questions on the CEQ to construct a variable referred to as the "extreme response index" defined by proportion of all responses in the questionnaire that are either 1's or 5's.

### 5.3.2 Different Fixed Effects used for Control

The different models were defined by the set of fixed effects used to control for the comparison across different courses and subject mixes as defined in Table 5.8. The frequency count of these variables is given in the Appendix F of the Resource Document for Australian resident and Australian citizen students who filled out the CEQ/GDS and are described in the histograms in Figure F.3.

<b>Model Fixed effects defined for the model</b>		<b>Appendix</b>
1	<i>Awdstrm1</i> Course as defined by University of Melbourne	F.2
2	<i>Duties</i> Duties at job since course completion	F.3
3	<i>Industry</i> Industry in which you work	F.4
4	<i>maj1</i> Major as defined by DEST codes (similar to <i>n_ceqmj</i> )	F.5
5	<i>mj1*mj_pc1</i> Department * the proportion of all subjects taken (1st)	F.6
	<i>mj2*mj_pc2</i> Department * the proportion of all subjects taken (2 <sup>nd</sup> )	
6	<i>n_ceqmj</i> Alternative DEST major code (similar to <i>maj1</i> )	F.7
7	<i>n_fac1*level</i> Faculty * level of the course combination	F.8

**Table 5.8:** The model definitions by the fixed effects accounted for (see appendix F in the Resource Document for this Report for details)

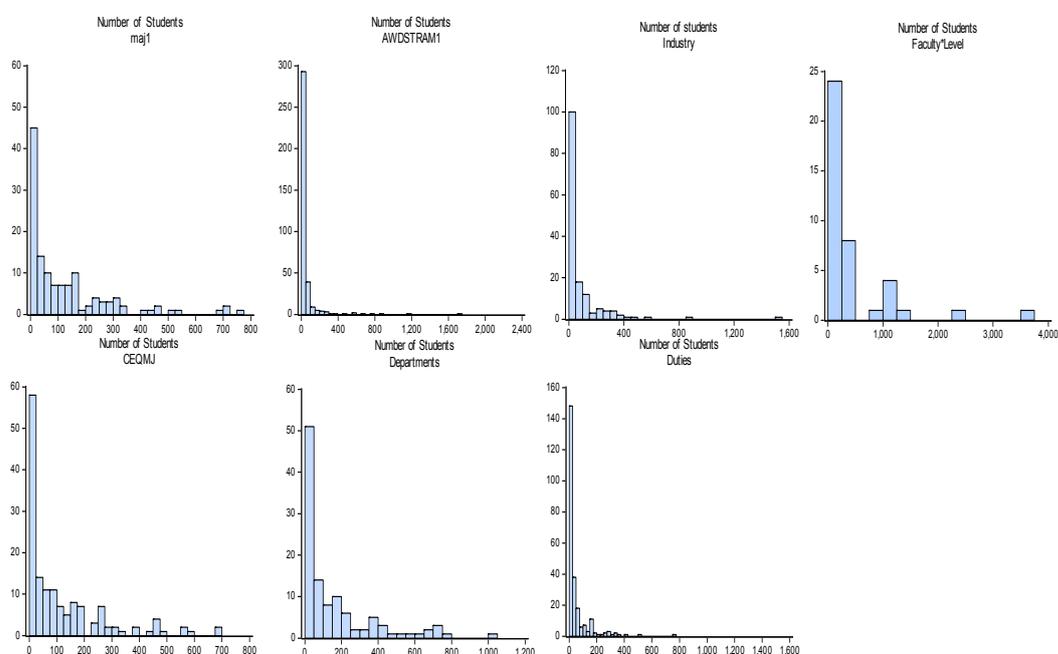
In model #1, we use the University of Melbourne course codes (*AWDSTRM1* see Table F.2). These were the first courses reported on the CEQ (to account for some students who completed multiple courses we use the first one on the form) as defined by the university and we found 363 in this sample. From this list, we noted that the Bachelor of Arts (BA) accounted for 10.15%, the Bachelor of Commerce (BComm) accounted for 6.71%, and the Bachelor of Science (BSc) for 4.98% of all the students who completed the CEQ at Melbourne in our study period. We found that the 10 largest courses had more than 45% of these students, whilst there 95 courses were listed with one to three students accounting for just 1.1% of the all the students in this sample. In sum, we found that these course definitions provided a very detailed description in only some cases, but an overly broad distinction for a much larger number of students.

Because the actual course names defined by *Awdstrm1*, as used in Model #1, may have provided little detail for the majority of students, we used a number of alternative categories to account for these fixed effects. In many cases these fixed



effects were similar in coverage to the course names, but they differed in the nature of their distribution as can be seen in Figure 5.4 which displays the histograms of the sizes of the categories as defined by each alternate definition of fixed effects that we employed. In Appendix F in the Resource Document for this report, we provide tables for each of these fixed effects definitions.

From Figure 5.4 it can be seen that the classifications of the major subjects studied variable definitions *maj1*, *department* and *ceqmj* resulted in the fewest number of very small groupings and just over 1000 in the largest department. The variables *duties* and *industry* were both taken from responses on the GDS of students' activities once completing their degrees. The major drawback to the use of these categories was the large number of responses (almost 25% of the sample) listed as *Unknown*. The models with the smallest number of groups were based on the faculty and degree level combination. There were around 70 non-zero combinations of which one (the Bachelor of Science) has almost 22% (recall that this sample includes only Australian resident students who completed the CEQ).



**Figure 5.4:** The histograms of the number of categories by size of the seven fixed effects (note that Duties and Industry exclude the Unknown category).

### 5.3.3 Interpretation of the Estimated Coefficients

Table 5.9 reports the coefficient estimates for the intercept and the continuous student experience variables as well as the *F*-statistics for the tests that dummy variables in each group are equal to zero. The dependent variable in this case was the Good Teaching Score (GTS) which is an average of the values for questions 101, 103, 115, 116, 110, and 127 as defined in Figure 5.2.

The continuous regressors can be sub-divided into two groups:

1. Those regressors that related to the individual who filled out the form such as: their age; average marks; the change in their marks in the last year from the



previous years; their average rank in the subjects they took during their last year; the number of courses for which they completed the survey; the number of subjects taken for their course; the number of hours they are currently working; and the log of their salary (plus 1 to account for zero).

- Those regressors that described the subjects the student took: the average of the subject enrolments; the difference between the enrolment when they took the subject and the average enrolment: the percentage of marks that were just under the next level (47, 48, 49, 64, 69, 74, 79): the average of the predicted QOT Question #2 based on the regressions for each department described above; the negative and positive deviations of the average of QOT Question #2 responses for the subjects taken, the average predicted QOT Question #2 score: and the deviation in the QOT Question #2 scores of the last 2 years as opposed to all the subjects taken.

	Effect	Ty	Model						
			1	2	3	4	5	6	7
1	AGE_Agebasedoncensusdate	CF	0.002881*	0.004527***	0.004939***	0.004009**	0.003868**	0.002974*	0.003958***
2	avg_lsm2_Avgq2byLSMEAN	CF	0.180542***	0.249703***	0.281099***	0.060578	0.07873	-0.14645***	0.326559***
3	avg_mrk_AvgMarkoverallsubjectstaken	CF	0.002449	0.002679*	0.002496*	0.003493**	0.002955*	0.004643***	0.00389***
4	c_mks_%ofClosemarksinsubjectstaken	CF	-0.00138**	-0.00135**	-0.00132**	-0.00162**	-0.00176**	-0.00175**	-0.00144**
5	d_mrk_Changeinavgmkforlastyear	CF	-0.00347***	-0.00318***	-0.0031***	-0.00304**	-0.00339***	-0.00277**	-0.0028**
6	d_qot_Changeinavgqotforlastyear	CF	0.01694	-0.00205	0.020355	0.013455	-0.0095	0.001205	0.017791
7	hours_WORKINGHOURS	CF	-0.00012	-0.00035	9.42E-06	-0.00002	-0.0001	0.00012	-0.00003
8	salary_Log(+1)ofstartingsalary	CF	0.002139	-0.00082	0.000442	0.002658	0.002902	0.002959*	0.002853
9	men_dev1_enrol-enr_avg	CF	0.00019*	0.000011	0.000024	0.000199*	0.000102	0.000216**	0.000083
10	menr_avg_AverageEnrolmentoverys	CF	-0.00021***	-0.00027***	-0.00033***	-0.00017**	-0.00018**	0.000013	-0.00019***
11	n_s_numberofsubjectstaken	CF	0.000255	0.000835	0.000741	0.000871	0.00129	0.001773***	0.001181*
12	nrv_dev2_diff<0avgqotq2	CF	0.382225***	0.549736***	0.495093***	0.405835***	0.50249***	0.29327**	0.432878***
13	pav_dev2_diff>0avgqotq2	CF	-0.01771	-0.01946	0.032588	0.012563	0.041911	0.152073	0.064957
14	mk1_Avgmkforlastyear	CF	0.002578***	0.002427***	0.002501***	0.002174***	0.002356***	0.002083***	0.002064***
15	CountrySEX	FS	0.651064	0.537277	0.469341	0.748808	0.60338	0.606922	0.497383
16	SURVYR_Yearurveyconducted	FS	0.326603	0.685733	0.874295	0.749992	0.786987	0.375465	0.931374
17	attendlfwork	FS	4.358575***	4.965811***	5.573659***	3.969243***	4.335034***	4.258501***	4.304005***
18	furtev_LEVELOFFurtherSTUDY	FS	3.196024***	3.135484***	2.701773***	3.805717***	3.237499***	3.354316***	2.515054***
19	level_LEVELOFQUAL	FS	4.584582***	13.07881***	10.64448***	12.98081***	23.87362***	19.76571***	19.930027
20	n_cour_numberofcoursesreportedon	FS	11.294179***	11.196***	10.45325***	9.236193***	8.485062***	6.471596***	8.210803***
21	Extremeresponseindex	RE	0.395139***	0.486829***	0.512436***	0.492315***	0.497142***	0.475554***	0.513163***
22	Extremeresponseindexsquared	RE	1.290638***	1.185695***	1.166777***	1.143996***	1.136349***	1.144269***	1.14589***

\*.1 < prob(t) < .05, \*\*.05 < prob(t) < .01, and \*\*\* prob(t) < .001

**Table 5.9:** The parameter estimates and the F-test parameters for the mixed models estimated for the Good Teaching Scale (GTS) using Australian Citizens and Permanent Resident.<sup>18</sup>

From Table 5.9 it can be seen that there was a uniform result showing older students (AGE in row 1) to be more positive their opinions of the teaching in their course. We also find that the higher their average mark and rank from their last year (mk1 in row 14), the more positive their reaction. However, if they have only recently started to receive higher marks, with their marks for their last year differing from the previous years (avg\_mrk in row 3), they were prone to evaluate more negatively. In sum, it seems that those students who did well throughout their course were positively disposed to the teaching, but that if they had only recently achieved higher marks, they tended to respond more negatively. Interestingly, the

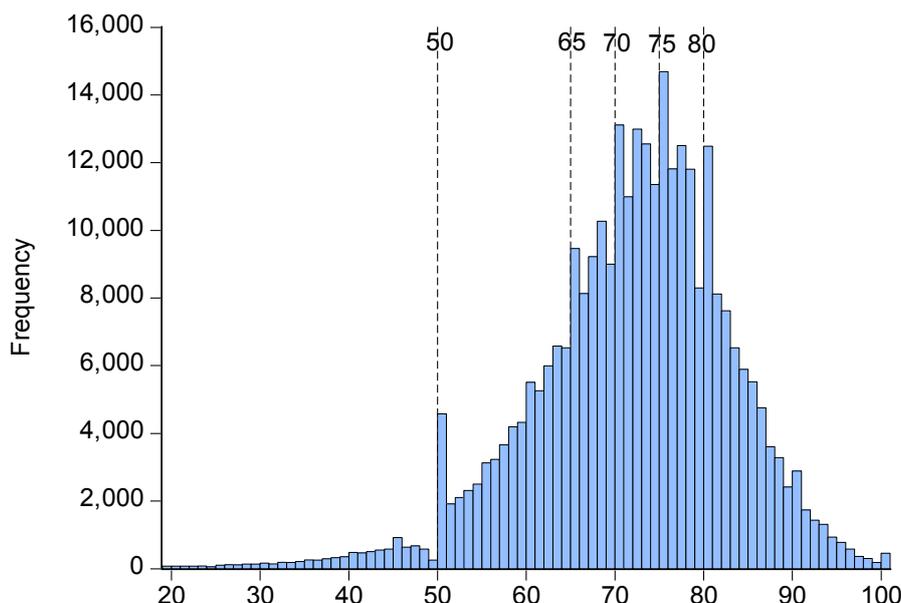
<sup>18</sup>CF = Estimated Coefficient, FS = F-statistic for the composite hypothesis that all the coefficients are equal to zero, and RE = the coefficients on the random effects.



number of hours in their jobs (*hours* in row 7) and the log of their salaries since they finished their degrees (*lsalary* in row 8) seemed to bear little weight in their reactions to their courses.

The next significant variable is the average percentage of the marks to all students in each subject that are close marks (*c\_mrks* in row 4, the proportion of marks that are just below the cut off for the next higher grade). From Table 5.9 we show that the higher the percentage of close marks, the lower the scores for good teaching – and this phenomenon appears in all the models specified. At The University of Melbourne the marks are given on a scale from 1 to 100. The categories are defined as: 50-64 is a pass, 65-69 is 3<sup>rd</sup> class honours or H3, 70-74 is a 2<sup>nd</sup> class honours division B or H2B, 75-79 as a 2<sup>nd</sup> class honours division A or a H2A, and 80-100 as a 1<sup>st</sup> class honours or H1. Thus obtaining a 49 would mean a failure. Some subjects are evaluated more closely to the cut-off values than others. In many cases a 79 would be rounded up to an 80 thus resulting in distributions of grades with a degree of lumpiness just above the cut offs.

Figure 5.5 is a histogram of all the marks received by the students who filled out the CEQ/GDS. From this figure, one can note that the marks just below the cut-off points (47, 48, 49, 64, 69, 74 and 79) are markedly lower due to the subject coordinator erring on the side of the student or what some researchers have referred to as leniency. However, some subject’s coordinators are more prone to ignore the proximity of the mark to the cut off values than others. This may be due to a number of circumstances such as grading by committees and where professional accreditation is connected with certain subjects. We find that the percentage of close grades varies by subject and that different students experience different proportion of subjects where close grades are issued. This measure can be used as a measure of the degree of “exactness” or lack of lenience used in grading by the subject coordinator. Interestingly, when a model was fit with the % of close grades earned by the particular student, it was not found to be a significant factor in any model.



**Figure 5.5:** The distribution of all marks issued to the students in this sample.



The next significant coefficient estimate from Table 5.9 is for the measure of the average value of the QOT Q2 “This subject was well taught” response which we have used after decomposing these scores. The observed QOT score for any subject can be decomposed into the expected level based on the subject  $k$   $\overline{QOT}_k$  and either a positive  $QOT_{ik}^+$  or negative  $QOT_{ik}^-$  result for the particular semester in which it was taught  $i$ . We can decompose the actual score observed from the subject teaching survey as:

$$\underbrace{QOT_{ik}}_{\text{Observed}} = \underbrace{\overline{QOT}_k}_{\text{Expected}} + \underbrace{QOT_{ik}^+}_{\text{Better}} + \underbrace{QOT_{ik}^-}_{\text{Worse}}$$

Where:

$$QOT_{ik}^+ = \begin{cases} (QOT_{ik} - \overline{QOT}_k) & \text{if } (QOT_{ik} - \overline{QOT}_k) > 0 \\ 0 & \text{if } (QOT_{ik} - \overline{QOT}_k) \leq 0 \end{cases}$$

and

$$QOT_{ik}^- = \begin{cases} 0 & \text{if } (QOT_{ik} - \overline{QOT}_k) > 0 \\ (QOT_{ik} - \overline{QOT}_k) & \text{if } (QOT_{ik} - \overline{QOT}_k) \leq 0 \end{cases}$$

We then compute the averages for each student. For example, the average Q2 for the subjects taken by a student who takes  $N$  subjects can be decomposed as:

$$\underbrace{\frac{1}{N} \sum QOT_{ik}}_{\text{Observed}} = \underbrace{\frac{1}{N} \sum \overline{QOT}_k}_{\substack{\text{Expectation} \\ \text{avg\_ism2}}} + \underbrace{\frac{1}{N} \sum QOT_{ik}^+}_{\substack{\text{Better} \\ \text{pav\_dev2}}} + \underbrace{\frac{1}{N} \sum QOT_{ik}^-}_{\substack{\text{Worse} \\ \text{nav\_dev2}}}$$

We define three variables for this decomposition *avg\_ism2* (in row 2) for the “expected value of Question #2 (Q2)” based on all the times the subject has been taught in the period of the QOT data, *pav\_dev2* (in row 13) is the average of the deviations of the QOT score for Q2 when the subject was taken by the student that are greater than the expected value, and *nav\_dev2* (in row 12) is the average of the negative deviations of the QOT score when the subject was taken by the student from the expected value for this subject (note these are all negative values).

For each department at The University of Melbourne, we computed a regression to model the Q2 response as a function of the subject, when it was taught, and the enrolment in the subject as described above in Section 5.2.2. We then predicted an average value for Q2 conditioning on the enrolment at the mean for the subject over the period of the data if there was sufficient data. Otherwise, we used the simple mean of the Q2 values registered for this subject. We identified this estimate as the “expected value of Q2” for the subject and we assumed that the student was aware of this prior to taking the subject. We also use observed Q2 as recorded for the subject for the particular semester and year when the student takes it. The difference between the expected or overall average Q2 and the Q2 for the subject in the particular semester and year studied was then computed. We then interpreted these two quantities as deviations from expectations. For subjects only appearing



once in the data these values are set to zero.

As shown in Table 5.9 (row 2) we record that the average expected Q2 scores (*avg\_lsm2* in row 2)) have a positive impact on the CEQ teaching responses (as defined by the GTS) for all but one of the models. From this finding we can conclude that, even when accounting for courses taken, the average Q2 scores had a significant positive impact on the CEQ response. Only in model #6 did we find that Q2 had a significant negative impact, there is no clear reason why the fixed effects used in this model should have resulted in this result.

Across all models we find an unambiguous result for the negative deviations from the average Q2 (*nav\_dev2* in row 12). If the Q2s for the subjects taken by the student are less than the expected Q2 for the subject, the impact is significant for every model. Note that, although this is estimated as positive, it is multiplied by a negative value thus the net effect is negative. This implies that, if the subject is not as well taught as one would expect, the student will score the course lower. In addition, the positive deviations (*pos\_dev2* in row 13) do not seem to matter – none of the positive deviations are significant. Doing better than expected does not lead to higher CEQ values for the GTS, only doing worse appears to have an impact.

The two categorical variables that were included as dummy variables to account for whether the individual was a full-time or part-time student and whether they worked proved to be a significant factor (the F-test for *attend\*fywork* in row 17). In particular, we found that students who were able to spend their last year in full-time study and not work provided the most positive responses. In addition, two of the categorical variables that defined their further study status were also found to be significant. In most of the models, we found that ex-students who were now studying for Masters and the Doctorate degrees (the F-test for *furlev* in row 18) had a positive impact on the survey result. Thus those students subsequently studying for graduate research degrees were more likely to provide positive feedback than those not pursuing a higher degree. The responses of students who pursued other degrees were not significantly different from those who were not undertaking further study. This may have been due to their empathy towards the teaching staff they encountered whilst enrolled in lower level courses.

The other set of coefficient estimates reported in Table 5.9 are for those characteristics of the subjects taken by the student. The average enrolments in subjects taken (*menr\_avg* in row 10, the average enrolment for each subject) had a significantly negative impact in all the models except #6 which uses the CEQ major codes as the fixed effects. This may be due to a small number of outlier subjects with particularly large enrolments. Again model #6 is the one case where the deviation from the normal enrolment (*men\_dev1* in row 9) had an impact, although it was positive which would indicate that larger enrolments for when the student took the class resulted in a more favorable impression. One explanation for this result could be that larger than average enrolments were indicative of subjects that were gaining in popularity. The number of subjects taken (*n\_s* in row 11) appears only to be a positive influence in model #6.



## 5.4 The student experience influence on the response to other CEQ scales.

The results described in Section 5.3 relate to the responses on the CEQ to the Good Teaching Scale (*GTS*) which is composed of the average response to questions in this chapter referred to as *CEQ101*, *CEQ103*, *CEQ115*, *CEQ116*, *CEQ110* and *CEQ127* as defined in Figure 5.2.<sup>19</sup> We also applied this analysis to the other scales of measure in the CEQ which are not directly related to the nature of instruction. The scales are *GSS* (Generic Skills), *CGS* (Clear Goals and Standards), *IMS*, (Intellectual Motivation), *LCS* (Learning Community), *LRS* (Learning Resources), and *OSI* (Overall Satisfaction). These scales are also defined in Figure 5.2.

<b>Dependent Variable</b>	<b>Model</b>						
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<i>CGS</i> <i>Clear Goals and Standards</i>	0.18	0.16	0.14	0.14	0.15	0.14	0.13
<i>IMS</i> <i>Intellectual Motivation</i>	0.21	0.20	0.19	0.19	0.20	0.21	0.18
<i>LRS</i> <i>Learning Resources</i>	0.21	0.18	0.17	0.18	0.19	0.19	0.17
<i>GSS</i> <i>Generic Skills</i>	0.23	0.21	0.20	0.20	0.21	0.22	0.19
<i>GTS</i> <i>Good Teaching</i>	0.26	0.24	0.23	0.24	0.24	0.26	0.22
<i>OSI</i> <i>Overall Satisfaction</i>	0.17	0.15	0.14	0.14	0.15	0.15	0.13

**Table 5.10:** R<sup>2</sup> for each model/dependent variable combination<sup>20</sup>.

Table 5.10 records the fit statistics for each model and dependent variable based on the square of the correlation between the fitted value and the actual values. From this table we note that model #1 performs as well if not better than all the other models, although they are all fairly similar in their performance. Note that model #7 has far fewer fixed effects than the other models which results in a lower degree of fit. The results for all the models are listed in Tables G.1 and H.1 in the Appendices G and H respectively.

To compare the estimated parameters for the models across the dimensions of the CEQ, we provide the corresponding coefficients and test statistics in Table 5.11, here we have listed the estimates as in Table 5.9 except, instead of across models for one dependent variable, we have listed the results across all six CEQ scales.

From Table 5.11 we note that when most of the coefficients in these models were significant they had a similar sign and magnitude to the model for the *GTS*. However, there were a few exceptions.

One primary difference was the coefficient on the average enrolment in the subjects taken (*menr\_avg* in row 10). For the *GTS* this coefficient was estimated as significantly negative. However, for the *LRS* (Learning Resources Score) this coefficient was estimated as positive. This might be an indication that larger class sizes are indicative of better resources. Note that the magnitude of both estimated parameters was almost equal but opposite in sign.

Another factor that appeared to influence results in different ways is the impact of age (*Age* in row 1). For the Generic Skills (*GSS*) and Overall Satisfaction Indicator (*OSI*) age had a negative impact on the response. However, the *GTS* coefficient

<sup>19</sup> Note in Chapter 4 we defined these items as *C1* to *C6* as presented in Table 4.2

<sup>20</sup> The complete set of results for these models is listed in Appendices G and H of the Resource Document for this report.



estimate for age is positive. For both GSS and OSI the magnitude of the effect was around double the positive impact on GTS.

Two factors that did not appear to have an influence on the GTS, but had a significant impact on other scores, are the average mark (*avg\_mrk* in row 3) earned over all subjects and the log of students' wages after they graduated (*lsalary* in row 8).

Effect	ty	CGS	GSS	GTS	IMS	LRS	OSI
1 AGE_Age based on census date	CF	0.001703	-0.00574***	0.002881*	0.002722	-0.00287*	-0.00408*
2 avg_lsm2_Avg Q2 by LSMEAN	CF	-0.10804**	0.031405	0.180542***	0.119377**	-0.00665	0.184715***
3 avg_mrk_Avg Mark over all subjects taken	CF	0.000678	0.001166	0.002449	0.00277*	0.002836**	0.003906**
4 c_mrks_% of Close marks in subjects taken	CF	-0.0008	0.000237	-0.00138**	-0.0001	-0.00081	0.000162
5 d_mark_Change in avg mark for last year	CF	-0.00317**	-0.00224**	-0.00347***	-0.00329**	0.000038	-0.00308**
6 d_qot_Change in avg qot for last year	CF	0.05449**	-0.00729	0.01694	0.003721	-0.01856	-0.00568
7 hours_WORKING HOURS	CF	-0.00014	0.000523	-0.00012	0.001268**	-0.00076	0.000578
8 lsalary_Log (+1) of starting salary	CF	0.003227*	0.004163**	0.002139	0.003038	0.006822***	0.006102***
9 men_dev1_enrol-enr_avg	CF	0.000162	0.00003	0.00019*	0.000064	-0.00003	-0.00003
10 menr_avg_Average Enrolment over yrs	CF	-0.00006	-0.00001	-0.00021***	-0.00004	0.000209***	-0.00009
11 n_s_number of subjects taken	CF	-0.00034	0.000806	0.000255	0.0001	0.00016	-0.00059
12 nav_dev2_diff < 0 avg qot Q2	CF	0.21634**	-0.02534	0.382225***	0.315805***	0.340942***	0.403102***
13 pav_dev2_diff > 0 avg qot Q2	CF	-0.0713	0.073336	-0.01771	-0.06734	-0.17533	-0.07389
14 mk1_Avg mk for last year	CF	0.000471	0.002303***	0.002578**	0.003474***	-0.00012	0.003026***
15 Country*SEX	FS	12.8743***	1.781679*	0.651064	0.917909	1.731831*	1.134225
16 SURYR_Year survey conducted	FS	9.070961***	2.572622*	0.326603	1.834845	1.053991	0.748619
17 attend*fywork	FS	57.16642***	3.124967***	4.358575***	3.352842***	2.830001**	4.701615***
18 furlev_LEVEL OF Further STUDY	FS	2.48454***	1.729066*	3.196024***	2.145381**	0.50712	1.533271
19 level_LEVEL OF QUAL	FS	1.571499	3.970604***	4.584582***	3.584497***	0.693196	1.995007*
20 n_cour_number of courses reported on	FS	9.30233***	1.604675	12.94179***	1.40218	2.022536	1.031647
21 Extreme response index	RE	-0.12266	1.457218***	0.395139***	1.990333***	1.11947***	1.34916***
22 Extreme response index squared	RE	1.098619***	-0.13901	1.290638***	-1.03621***	0.20225**	-

\*.1 < prob(t) < .05, \*\*.05 < prob(t) < .01, and \*\*\* prob(t) < .001

**Table 5.11:** Coefficient estimates across the dependent variables using model #1 which uses the course code from the University of Melbourne for the definition of the fixed effects.<sup>21</sup>

The log of wages was estimated to have a positive impact on CGS, GSS, LRS and OSI. Thus, given hindsight from within a working position, these graduates had a better appreciation of these aspects of the institution. Note that this regressor had a number of zero values (the log of wages has been scaled to have zero as the minimum value) for those graduates who may not be looking for work or had not found a position but were actually pursuing a higher degree in another course of study.

The average mark earned by the graduates (*avg\_mrk* in row 3) was another regressor that showed a significant coefficient in determining other effects instead of the GTS. In this case, we found that IMS, LRS and OSI values were all positively influenced by higher average marks. This effect was highest for the influence on the Overall Satisfaction Score (OSI).

## 5.5 Discussion

This chapter has examined the factors that influence the responses the Course

<sup>21</sup> As in Table 5.8, CF = Estimated Coefficient, FS = F-statistic for the composite hypothesis that all the coefficients are equal to zero, and RE = the coefficients on the random effects.



Experience Questionnaire that can be identified from student study experiences. We find that both marks received and the measures of teaching quality attribute to these factors. We also confirm the findings from earlier studies that the marks a student receives can influence their responses to evaluation surveys. We have also been able to identify the influence of the student's rank as well. In particular, we find that the students' rank in the final year of their studies has a significant impact on their perception of their course of study and subsequently on their survey responses. Conversely the impact of the students' ranking from earlier years was found to be negligible. Influences generated by their average marks were also positive, but less strong. Moreover, if the students' marks changed in the course of their last year of study, they were less prone to evaluate teaching positively. In general, we may conclude that marks do not have an unambiguous positive impact. It was also found that the student's average mark in the last year was not a significant factor when the average mark and the change in their average mark in the last year was replaced by the average from the last year. This seems to bear out earlier research that suggests that marks alone are not the prime factor in considering survey responses once the characteristics of the degree are established.

The other major finding from these estimated models is that the impact of the average responses on the subject specific teaching survey (as measured by Q2 on the QOT) can be mitigated when we account for the department in which the subjects were taken. If we include the department as a fixed effect (the case in models #4 and #5) we find that the average QOT response on Q2 is not a significant factor in the GTS responses. This would agree with our findings that the subject specific evaluations are strongly influenced by the department in which the subject is taught.

However, we discovered an asymmetric influence of how subject specific average Q2 on the QOT deviated from the expectation for Q2 responses in the subject. There was an unambiguous negative influence of a lower than expected QOT while there was no corresponding positive bounce for a greater than expected QOT. Thus, when a subject achieves a higher than expected QOT it does not add to the course evaluation. The negative finding for positive deviations from expectations for subject are quite concerning for the use of subject measures from the surveys conducted in classes. The strong findings for negative deviations imply that variations in subject evaluations do not "even out".

We find that students have a negative reaction to having a sharp distinctions made on their marks. The negative reaction to subjects where students do not benefit by degrees of doubt or leniency indicates that it may be necessary to acknowledge that the evaluation processes or marking styles have a bearing on student's perceptions of a course. We also find evidence that this measure of leniency or lack of, may have acted as a proxy for the manner in which a subject is conducted since we also discovered that the aversion to "close marks" was based on the subject average and not the experience of the particular student whose response was being modeled.

In summary, we have found a number of positive influences on the Good Teaching Scale (GTS). These include:

- The older the student
- The higher their average marks over the entire set of subjects
- The higher their average rank in subjects taken in the last year
- The higher the average response to Q2 on QOTs in subjects taken



- If going on for further study
- If did not work and attended full-time in last year.

In contrast we were also able to determine a number of negative influences as:

- If marks changed in the last year
- The more other courses they completed
- The greater the average enrolments in classes they took
- The greater the proportion of “close marks” in the subjects taken
- If the Q2 in the subjects they took was lower than the average for the subject.

In our examination of the other dimensions of the student experience measured by the CEQ, we found that some factors which influenced the Good Teaching Scale were not found to be important factors in the determination of these other scores. In the case of the two most prominently used other CEQ scores, we even found evidence that, in opposition to the model of the GTS, that the age of the student was a negative factor in evaluations by way of the Generic Skills Score (GSS) and the Overall Satisfaction Index (OSI). These GSS findings may be an indication that older students would more likely to have already acquired a set of Generic Skills and had a lower appreciation of these skills provided by the University.

The other major finding which contrasted with our conclusions concerning the Good Teaching Scale was that class sizes have a positive impact on the score for the evaluation of Learning Resources which is the inverse of all our other findings. Again we may conclude that this was due to the better facilities that could be found in courses that have larger numbers of students and this might be an argument for the existence of economies of scale when evaluating Learning Resources.



## Chapter 6. Discussion

At the outset of this study we aspired to establish how factors signaled by the results of a typical subject specific teaching quality instrument (TQI) at a tertiary institution, might be translated into graduates' responses in the Course Evaluation Questionnaire (CEQ). We discovered that the highly varied nature of the TQIs in Australian Universities meant that the direct translation of the responses on local TQIs to responses on the national CEQ requires a specialised mapping for every case. To solve this problem we defined a classification of the various TQIs and used this classification scheme to construct a new composite TQI. The rationale for the development of this new TQI was to account for problems in matching the CEQ created by possible idiosyncratic characteristics of the local surveys.

We investigated the relationship between the new TQI, the local TQI and the "Good Teaching Score" questions from the CEQ by administering a composite survey that included questions from all three, to students taking subjects at the four participating institutions. The analysis of these survey responses found that, although there were some similarities in the responses to the CEQ questions and the new TQI items, the correspondence between the CEQ and the local TQIs was found to be much lower. In other words, the responses on the local TQI were a poor indicator of the information elicited by the teaching related questions from the CEQ on the same survey. After applying a number of methods for comparing these survey responses we also found little evidence of a direct correspondence between the new TQI measures and the CEQ. We do not find this to be a surprising result due to the different design intensions of the surveys on which it is based. However, this may be one reason for including similar questions to the CEQ on the TQIs. Based on this finding we then pursued another course of investigation where past average subject TQI results would be matched to the responses on subsequent CEQ surveys and conditioning our inferences on additional subject and student information.

Our survey of the literature that discussed factors that influence student responses to TQIs found extensive evidence that factors, other than those under the control of the instructor and institution, were often critical factors in the determination of average TQI outcomes. We refer to this literature as an indication that it is necessary to condition the CEQ response by factors that would not be measured by the past average TQI responses for the subjects taken by the student. Thus, to establish a statistical relationship between the average TQI, as measured in individual subjects, and the subsequent responses to their course evaluation surveys, we collected additional information for each student. This information is in addition to the data collected on the CEQ and on the companion Graduate Destination Survey (GDS) and is based on the transcripts of the responding students.

Of the four participating universities we discovered that we were only able to augment the CEQ/GDS data with data from the University of Melbourne. In this case we were able to combine both the CEQ/GDS responses with the student characteristics, their specific marks in the subjects they took, the characteristics of the subjects they took as to mark distribution and enrollments, and the average TQI scores for the subject as recorded on the internal survey. This combined data set enabled us to estimate a number of models to establish the degree to which individual CEQ responses were influenced by measures of teaching quality that



could be derived from the average TQI responses recorded in the subjects taken by the student cohort. As a byproduct of these models we discovered that a number of other characteristics of the subjects, and of the students, influenced graduates' perceptions of their courses.

Our results were found to be robust to various strategies for classifying graduates by: the subjects the students took, the courses they completed, the Faculties in which they were enrolled, the departments in which they took the majority of their subjects, what degree level they completed, and their circumstances after graduation. Combinations of the inclusion of conditioning variables based on these distinctions allowed us to define a series of models that we estimated using a mixed regression estimation procedure. In addition, this procedure allowed us to include random effects based on the response style used by the student in completing the form.

In all these models, we found the consistent result that students were negatively influenced when the subjects they took were rated lower on the TQIs administered in the subjects they took than how they may have anticipated how the subject was usually taught. However, we also observed that the opposite case, when the TQI are higher than expected, resulted in no appreciable increase in the CEQ response. From this we conclude that there is an inherent asymmetry in the responses by students to lower subject evaluations than to higher ones.

Another characteristic found to be important was the influence of the type of marking in the subjects taken by the respondent. In particular, the impact of exactness or non-lenience in the marking of the subject was found to have a negative influence on the course evaluation. Whether this is due to: the empathy of the respondent to other students in the subject, a characteristic of the subject content (i.e. the subject formed part of a professional accreditation or is used as a filter for further study), or a characteristic of the instructor that cannot be determined from the information available. We verified that this was not due to personal experience when we found that the experiences by the respondent of "close marks" were not influential in the student's responses.

We also discovered that the response rate to the CEQ was not random. We found that the proportion of the students in a given subject, who subsequently completed the form on graduation, could be partially predicted by characteristics of the subject. In addition, we also discovered that the characteristics of the subject that had the greatest impact on increasing response were also the characteristics of subjects that generated negative responses on the CEQ by students that experienced them. This indicates that a higher proportion of students that experienced subjects with these adverse conditions responded to the CEQ. In other words, those students who were exposed to subjects that have less desirable characteristics have responded more readily. By including these characteristics in our model we have, in part, been able to account for the influence of the subjects taken on the scores, however, modelling the CEQ scores without conditioning them on the subject characteristics may lead to biased inferences.

One objective of this study was to map a typical process by which inferences could be drawn from the TQI to the subsequent CEQ for any institution. Unfortunately, in this quest we have fallen short due to the complexity of local TQIs and the policies by which they are implemented. We discovered that in many institutions (3 out of the 4 in this study as well) there is no history of complete TQI surveys throughout a



University or Faculty or even a Department that can be used to construct a description of the student's experiences while completing their degree. We also found that in many instances surveys may not be conducted unless the instructor is to be considered for promotion. Due to these difficulties we were only able to create a history of subjects based on the average TQIs for a particular student for students enrolled at the University of Melbourne where due to its universal application and static nature, it was comparable over students in different degree programs and over time.<sup>22</sup> In this regard we have discovered that the University of Melbourne is fairly rare and possibly unique in Australia.

We have endeavored to document the steps we have taken in our research in order to provide sufficient detail to enable others to replicate or to use aspects of our research. In this regard we feel that the new survey that we trialed at the four institutions can be used to improve TQIs in use. Although only including questions on the TQI that mirrors the CEQ does not account for the alternative factors that we found to be important in determining the ultimate CEQ responses. We found that such aspects of a subject such as leniency or lack thereof in assessment is also a factor in the subsequent course evaluation. In addition, we discovered that the subjects and the departments in which the majority of subjects were taken are important factors in determining the CEQ responses thus indicating that comparisons of departments and faculties should not be conducted on the raw CEQ averages and need to be conditioned by the characteristics of these cases.

Future research in this area should consider the development of better sampling techniques. We have found that by accounting for a number of student characteristics that are not currently on the CEQ we can improve the quality of the information gained from post graduation surveys. In this study we have attempted to add as much detail as possible for each student's response but the quality of these inferences still relies on the assumption that the sample used in estimation is representative for those factors that we cannot measure. However, the evidence that sample selection bias may be present in the current sample indicates that it is necessary to strive for quality over quantity in the sample used to draw inferences from the CEQ. A better sampling method that employs such methods as in-person interviews may be necessary to overcome the difficulties encountered with the present survey method.

---

<sup>22</sup> Alas this aspect of the TOQ at Melbourne is about to be lost due to the move in the first semester of 2011 to a different survey that is now administered electronically.



## References

- Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not know. *Journal of Educational Psychology, 82*(2), 219–231.
- Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research, 52*(3), 446–464.
- Access Economics (2005) *Review of Higher Education Outcome Performance Indicators Report for Department of Education, Science and Training*, April Australian Government Printers
- Aigner, D. J., & Thum, F. D. (1986). On student evaluation of teaching ability. *Economic Education, 17*(4), 243–265.
- Athhiyaman, A. (2001). Graduates' Perception about Business Education: An Exploratory Approach. *Journal of Further and Higher Education, 25*(1), 5–19.
- Ballantyne, R., Bain, J. D., & Packer, J. (1999). Researching university teaching in Australia: Themes and issues in academics' reflections. *Studies in Higher Education, 24*(2), 237–257.
- Barrie, S. and P. Ginns (2007). The Linking of National Teaching Performance Indicators to Improvements in Teaching and Learning in Classrooms. *Quality in Higher Education, 13*, 275–286.
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles: A Journal of Research, 43*(5/6), 407–417.
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology, 79*(3), 308–314.
- Bedgood, R. E., & Pollard, R. J. (1999). Uses and misuses of student opinion surveys in eight Australian universities. *Australian Journal of Education, 43*(2), 129–156.
- Boice, R. (n.d.). Countering common misbeliefs about student evaluations of teaching. from [http://www.missouri.edu/~petwww/evaluation/student\\_eval/common\\_misbeliefs.htm](http://www.missouri.edu/~petwww/evaluation/student_eval/common_misbeliefs.htm)
- Bond, T. and C. Fox (2007), *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*, 2nd Edition, Lawrence Erlbaum Associates Inc.
- Bosshardt, W., & Watts, M. (2001). Comparing student and instructor evaluations of teaching. *Journal of Economic Education, 32*(1), 3–17.
- Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work: Enhancing individual and institutional performance. San Francisco: Jossey-Bass.
- Broomfield, D., & Bligh, J. (1998). An Evaluation of the Short Form Course Experience Questionnaire with Medical Students. *Medical Education, 32*, 367–369.
- Brown, G., & Atkins, M. (1993). *Effective teaching in higher education*. London: Routledge.



- Byrne, M., & Flood, B. (2003). Assessing the Teaching Quality of Accounting Programs: an Evaluation of the Course Experience Questionnaire. *Assessment & Evaluation in Higher Education*, 28(2), 135–145.
- Casey, R. J., Gentile, P., & Bigger, S. (1997). Teaching appraisal in higher education: An Australian perspective. *Higher Education*, 34(4), 459–482.
- Cashin, W. E. (1988). Student ratings of teaching: A summary of the research. *IDEA Paper No. 20*. Kansas.
- Cashin, W. E. (1990). Students do rate different academic fields differently. In M. Theall & J. Franklin (Eds.), *Students ratings of instruction: Issues for improving practice: New directions for teaching and learning* (pp. 113–121). San Francisco: Jossey-Bass.
- Cashin, W. E. (1995). Student ratings of teaching: The research revisited. *IDEA Paper No. 32*, 32(Sept).
- Cashin, W. E., & Downey, R. G. (1992). Using global student ratings items for summative evaluation. *Journal of Educational Psychology*, 84(4), 563–572.
- Centra, J. A. (1993). Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass.
- Costin, F., Greenough, W. H., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 8(511–35).
- Cuseo, J. (2002). Evaluating new-student seminars and other first-year courses via course-evaluation surveys: Research-based recommendations regarding instrument construction and administration, data analysis, data summary, and reporting results. <http://www.brevard.edu/fyc/fya/CuseoLink.htm>
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198–1208.
- Davies, M., Hirschberg, J., Johnston, C., & Lye, J. (2006). 'What Influences Teaching Evaluations? Evidence from a Major Australian University. *The Business Review*, 6(1), 146–152.
- Davies, M., Hirschberg, J., Johnston, C., & Lye, J. (2007). *An Exploration of the Relationship between the Quality of Teaching Survey and the Course Experience Questionnaire*. Paper presented at the Quantitative Analysis of Teaching and Learning in Higher Education in Economics, Commerce and Business, The University of Melbourne.
- Davies, M., Hirschberg, J., Johnston, C., & Lye, J. (2007). Systemic Influences on Teaching Evaluations: The Case for Caution. *Australian Economic Papers*, 46(1), 18–38.
- Davies, M., Hirschberg, J., Johnston, C., and Lye, J., (2007). An Analysis of the Questions on University Teaching Surveys and the Universities that Use Them: the Australian Experience, Department of Economics, University of Melbourne, Working Paper, No. 994.
- Davies, M., Hirschberg, J., Johnston, C., Lye, J., & McDonald, I. (2005, 13–15 June, 2005). *Is it Your Fault? Systemic Influences on Student Evaluation of Teaching in Tertiary Institutions*. Paper presented at the First International Conference on Enhancing Teaching and Learning through Assessment (Revise and resubmit *Assessment and Evaluation in Higher Education*).



- Davies, M., Hirschberg, J., Lye, J. and C. Johnston "A Systematic Analysis of Quality of Teaching Surveys", 2010, *Assessment & Evaluation in Higher Education*, Vol. 35, No. 1, 87–100.
- Department of Employment, Science and Training (2004). *Our Universities: Backing Australia's future — The Higher Education Reforms*, Canberra, Australian Government printer.
- DEST (2006). Learning and Performance Fund 2006 Administrative Information for Providers, downloaded from <http://tinyurl.com/2ygzxg> 12<sup>th</sup> November 2007.
- Downie, N., & Moller, I. (2005). The Ramsden Course Experience Questionnaire: A Pilot Study of Final-Year Students taking Hospitality, Leisure, Sport and Tourism Degrees. *Journal of Hospitality, Leisure, Sport and Tourism Education*, 1(1).
- Dwinell, P. L., & Higbee, J. L. (1993). Students' perceptions of the value of teaching evaluations. *Perceptual and Motor Skills*, 76, 995–1000.
- Eley, M. (2001). The Course Experience Questionnaire: Altering Question Format and Phrasing Could Improve the CEQ's Effectiveness. *Higher Education Research and Development*, 20(3), 293–312.
- Evered, R., (2008), *The Good Universities Guide 2008*, Melbourne: Hobsons Australia
- Everitt, B., (1974), *Cluster Analysis*, Heinemann Educational Books Ltd, London.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9, 199–242.
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10(2), 149–172.
- Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, 21(1), 45–116.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students colleagues, administrators and external (neutral) observers. *Research in Higher Education*, 30(2), 137–194.
- Feldman, K. A. (1992). College students' views of male and female college teachers: Part 1 — evidence from the social laboratory and experiments. *Research in Higher Education*, 33, 317–375.
- Feldman, K. A. (1993). College students' views of male and female college teachers: Part ii — evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34(2), 151–211.
- Felton, J. Mitchell, J & Stinson, M. (2004). Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness , *Assessment & Evaluation in Higher Education*, 29 (1) Feb 92–107.
- Fries, C. J., & McNinch, R. J. (2003). Signed versus unsigned student evaluations of teaching: A comparison. *Teaching Sociology*, 31(3), 333–344.
- GCCA(1999). Institutional Arrangements for Student Feedback Project Report. A report of the investigation into university practices for obtaining student feedback. Parkville, Victoria, Graduate Careers council of Australia Ltd.



- Gokcekus, O. (2000). How do university students value economics courses? A hedonic approach. *Applied Economics Letters*, 7, 493–496.
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182–1186.
- Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4), 743–751.
- Griffin, B. W. (2000). Instructor reputation and student educational psychology. from [http://coe.gasou.edu/foundations/bwgriffin/research/instructor\\_reputation\\_2000.htm](http://coe.gasou.edu/foundations/bwgriffin/research/instructor_reputation_2000.htm)
- Griffin, P., Coates, H.; Mcinnis, C. and R. James (2003), “The Development of an Extended Course Experience Questionnaire”, *Quality in Higher Education*, 9, 259–266.
- Grimes, P. W., Millea, M. J., & Woodruff, T. W. (2004). Grades—who's to blame? Student evaluation of teaching and locus of control. *Journal of Economic Education*, 35(2), 129–147.
- Hativa, N. (1998). Lack of clarity in university teaching: A case study. *Higher Education*, 36(3), 353–381.
- Hativa, N., & Raviv, A. (1993). Using a single score for summative evaluation by students. *Research in Higher Education*, 34(5), 625–646.
- Haynes, B. (2002). *The use of student evaluations as an indicator of teaching quality in higher education*. Paper presented at the Conference of Australian Association for Research in Education, Brisbane.
- Hepworth, D., & Oviatt, B. E. (1985). Using student course evaluations: Findings, issues, and recommendations. *Journal of Social Work Education*, 21(3), 105–112.
- Howard, G.S. & Maxwell S.E. (1982) Do grades contaminate student evaluations of instructions? *Research in higher education*, 18(2), 175–188.
- Husbands, C. T. (1996). Variations in students' evaluations of teachers' lecturing and small-group teaching: A study at the London school of economics and political science. *Studies in Higher Education*, 21(2), 187-206.
- Husbands, C. T. (1997). Variations in student evaluations of teachers' lecturing in different courses on which they lecture: A study at the London School of Economics and political science. *Higher Education*, 33(1), 51–70.
- Husbands, C. T., & Fosh, P. (1993). Students' evaluation of teaching in higher education: Experiences from four European countries and some implications of the practice [1]. *Assessment & Evaluation in Higher Education*, 18(2), 95–114.
- Jackson, D. L.; Teal, C. R.; Raines, S. J.; Nansel, T. R.; Burdsal, C. S.; Force, R. C. (1999). The Dimensions of Students' Perceptions of Teaching Effectiveness. *Educational and Psychological Measurement*; 59, 580–96.
- Kaufman, L., and Rousseeuw, P. J. (1990). *Finding Groups in Data*. New York: John Wiley & Sons.
- Kolitch, E., & Dean, A. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education*, 24(1), 27–42.



- Lawless, C. J., & Richardson, J. T. E. (2002). Approaches to Studying and Perceptions of Academic Quality in Distance Education. *Higher Education*, 44, 257–282.
- Liaw, S. H., & Goh, K. L. (2002). Biases in student evaluation of teaching: The case of faculty of economics & administration university of Malaya. *FEA Working Paper, No. 2002–3*.
- Lizzio, A., Wilson, K., & Simons, R. (2002). University student's perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education*, 27(1).
- Low, G. (1999). What respondents do with questionnaires: Accounting for incongruity and fluidity. *Applied Linguistics*, 20(4), 503–533.
- Lowman, J., & Mathie, V. A. (1993). What should graduate teaching assistants know about teaching? *Teaching of Psychology*, 20(2), 84–88.
- 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), 339–352.
- Marsh, H. W. (1982). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement*, 6, 47–59.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253–388.
- Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education*, 64(1), 1–18.
- Marsh, H. W., & Dunkin, M. (1992). Students' evaluations of university teaching: A multi-dimensional perspective. In J. C. Smart (Ed.), *Higher education: Handbook on theory and research* (pp. 143–234). New York: Agathon Press.
- Marsh, H. W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education*, 7, 300–314.
- Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217–251.
- Marsh, H. W., & Ware, J. E., Jr. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of Educational Psychology*, 74(1), 126–134.
- McCormick, C. (1996). A focus group approach to undergraduate psychology program assessment. Paper presented at the Annual Meeting of the Southeastern Psychological Association (Norfolk, VA, March, 1996).
- McKeachie, W. J. (1978). *Teaching tips: A guidebook for the beginning college teacher*. Lexington, MA: Heath.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe: Bulletin of the AAUP*, 65(6), 384–397.



- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218–1225.
- McKeachie, W. J., & Lin, Y. G. (1979). A note on validity of student ratings of teaching. *Educational Research Quarterly*, 4(3), 45–47.
- Mehdizadeh, M. (1990). Loglinear models and student course evaluations. *Journal of Economic Education*, 21(1), 7–21.
- Millea, M., & Grimes, P. W. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, 36(4), 582–590.
- Morrison, D. F., (1976), *Multivariate Statistical Methods*, McGraw-Hill, New York, NY.
- Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2), 250–261.
- Nelson, J.P., & Lynch, K.A., (1984) Grade inflation, Real income, Simultaneity and teaching Evaluations . *Journal of Economic Education*, 15(1), 21–37
- Neumann, R. (2000). Communicating student evaluation of teaching results: Rating interpretation guides (rigs). *Assessment & Evaluation in Higher Education*, 25(2), 121–134.
- Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 72(3), 321–325.
- Patrick, J., & Smart, R. M. (1998). An empirical investigation of teacher effectiveness: The emergence of three critical factors. *Assessment & Evaluation in Higher Education*, 23(2), 165–178.
- Pozo-Munoz, C., Reboloso-Pacheco, E., & Fernandez-Ramirez, B. (2000). The 'ideal teacher'. Implications for student evaluation of teacher effectiveness. *Assessment & Evaluation in Higher Education*, 25(3), 253–263.
- Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology*, 135(3), 259–268.
- Ramsden, P. (1991a). A Performance Indicator of Teaching Quality in Higher Education: the Course Experience Questionnaire. *Studies in Higher Education*, 16(2), 129–150.
- Ramsden, P. (1991b). Report on the CEQ Trial. in R. Linke (Ed.), *Performance Indicators in Higher Education* (Vol. 2). Canberra: Australian Government Publishing Service.
- Ramsden, P. (2003). *Learning to Teach in Higher Education*. New York: Routledge Falmer, Second Edition.
- Ramsden, P., & Entwistle, N. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology*, 51, 368–383.
- 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. (2003). Approaches to Studying and Perceptions of Academic Quality in a Short Web-Based Course. *British Journal of Educational Technology*, 34(4), 433–442.



- Russell, P. and Rao, T. (1940). On habitat and association of species of anopheline larvae in south-eastern Madras. *Journal of the Malaria Institute of India*, 3, 153–178.
- Sadlo, G., & Richardson, J. T. E. (2003). Approaches to Studying and Perceptions of the Academic Environment in Students Following Problem-Based and Subject-Based Curricula. *Higher Education Research and Development*, 22(3), 253–274.
- Saroyan, A., & Snell, L. (1997). Variations in lecturing styles. *Higher Education*, 33(1), 85–104.
- Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education*, 39(3), 40.
- Sheehan, D. S. (1975). On the invalidity of student ratings for administrative personnel decisions. *Journal of Higher Education*, 46(6), 687–700.
- Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25(4), 397–405.
- 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.
- Sixbury, G. R., & Cashin, W. E. (1995). Comparative data by academic field. In *Idea technical report, vol. 10*: Center for Faculty Evaluation and Development, Manhattan, Kansas: Kansas State University.
- Smith, G. S. (2004). Assessment strategies: What is being measured in student course evaluations? *Accounting Education*, 13(1), 3–28.
- Solas, J. (1990). Effective teaching as construed by social work students. *Journal of Social Work Education*, 26, 145–154.
- Ting, K. (2000). A multi-level perspective on student ratings of instruction. *Research in Higher Education*, 41(5), 637–661.
- Trigwell, K., & Prosser, M. (1991). Improving the Quality of Student Learning: The Influence of the Learning Context and Student Approaches to Learning on Learning Outcomes. *Higher Education*, 22, 251–266.
- Van Herk, Hester, Ype H Poortigna and Theo M. M. Verhallen, 2004, "Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries", *Journal of Cross-Cultural Psychology*, 35, 346–360.
- Wachtel, H. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191–211.
- Wagner, Z. M. (1999). Using student journals for course evaluation. *Assessment & Evaluation in Higher Education*, 24(3), 261–272.
- Williams, R. (2008), "Ranking Australian Universities: Controlling for Scope", *Higher Education in Europe*, 33, pages 331 – 344.
- Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" problems with student ratings of instructors and courses. *Change*, 29(5), 12–23.
- Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education* (16 January, 1998), A12.



- Wolfer, T. A., & Johnson, M. M. (2003). Re-evaluating student evaluation of teaching: The teaching evaluation form. *Journal of Social Work Education, 39*(1), 111–121.
- Worthington, A. C. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assessment & Evaluation in Higher Education, 27*(1), 49–64.
- Zangenehzadeh, H. (1988) Grade inflation: A way out. *Journal of Economic Education 19*(3), 217–230.





**AUSTRALIAN  
LEARNING  
& TEACHING  
COUNCIL**

Promoting excellence in higher education

PO Box 2375 Strawberry Hills NSW 2012 Australia  
Telephone 02 8667 8500 Facsimile 02 8667 8515  
[www.altc.edu.au](http://www.altc.edu.au)  
ABN 30 109 826 628