The influence of student experiences on post-graduation surveys

Joe Hirschberg and Jenny Lye

Department of Economics,

University of Melbourne,

Victoria 3010, Australia

j.hirschberg@unimelb.edu.au

jnlye@unimelb.edu.au

Corresponding author: Joe Hirschberg

December, 2014

Key words: Course Experience Questionnaire, teaching quality instruments, students' perceptions, grade distributions, class size.

Abstract

This study attempts to establish the extent to which in-class *Teaching Quality Instruments* can be used to predict post-graduation survey results. It examines the responses for the *Good Teaching Scale* of the *Course Experience Questionnaire* (*CEQ*) administered to 10,433 students who completed their studies at a major Australian tertiary institution from 2003 to 2005 using a unique data set that matched student records and measures of class characteristics to the individual survey responses.

The findings indicate that the overall degree experiences of particular students can be predicted by measures of class differences as measured by *Teaching Quality Instruments* and the grade distributions of the classes they completed. These factors are in addition to the effects of student's own performance as measured by their grades, their field of study and their post-graduation experience.

It was found that in-class administered *Teaching Quality Instrument*s have an asymmetric influence on post-graduation survey results. Higher than expected *Teaching Quality Instrument* scores appear to have little impact, however, lower than expected results were found to have a significant negative impact on post-graduation recollections. Additionally, the grade distribution in classes taken was also found to be an important factor in explaining variation in degree satisfaction.

Introduction

Increasingly, quality assurance in higher education has emphasised quantitative measures of institutional performance. Since 1993 the Course Experience Questionnaire (*CEQ*) has been administered to all recently graduating students in Australia approximately four months after the completion of their studies (Ramsden, 1991a, 1991b; Ramsden and Entwistle, 1981). Its scales have been found to possess a number of useful properties (Eley, 2001; Richardson, 1994; Trigwell and Prosser, 1991). Originally, the results were confidentially provided to universities, although summaries have appeared in third-party reviews of tertiary institutions such as *The Good Universities Guide*.

In 2006, the Australian government introduced the *Learning and Teaching Performance Fund* model to reward higher education providers that best demonstrated excellence in teaching and learning for undergraduate domestic students. The cross-institutional comparisons of *CEQ* featured as a primary indicator for this funding model. In 2011, a new version of the survey entitled the *University Experience Survey* was proposed to evaluate Australian higher education (Radloff et al, 2012). This new survey incorporated the section of the *CEQ* referred to as the *Good Teaching Scale* examined in the analysis described here.

One limitation of surveys such as the *CEQ* is their inability to measure student perceptions of individual lecturers, classes taught and units of study. In addition, the retrospective and aggregate nature of these results has made it difficult for institutions to use the *CEQ* results in a timely manner to target internal quality improvement activities. To perform these functions, all Australian tertiary institutions have developed their own survey instruments to measure student perceptions of individual instructors and units of study (see Davies et al., 2010, for an analysis of these instruments). *Teaching Quality Instruments*, with names such as *QOT*s, *SET*s, *LET*s and *TEVAL*s, are a major element in the evaluation of Australian higher education and are useful for measuring differences between individual lecturers and the performance of separate teaching units within the institution.

Virtually all institutions have relied on these *Teaching Quality Instruments* to construct regular annual performance measures in relation to teaching quality and student satisfaction.

Thus, although both the *CEQ* and the local *Teaching Quality Instrument* measures have played a role in higher education quality assurance, little is known about the relationships between the responses to these instruments by the same cohort of students. Whereas the *CEQ* was intended to evaluate students' perceptions of their degrees, the *Teaching Quality Instrument*s have been designed to measure their perceptions of teaching at the class and department level. For this reason, one could assume that the quality of teaching element of the *CEQ* provides a composite measure of the teaching experience and thus the *CEQ* and *Teaching Quality Instrument* measures are related. Prior to the analysis described here, this hypothesis had not been formally tested. There is evidence that *CEQ* results are systematically influenced by a range of factors (Graduate Careers Australia, 2009). The research described here employs data from a major Australian university to examine the factors that influence responses to the *CEQ* with particular attention to the results of *Teaching Quality Instruments* conducted in the classes taken by the respondents as well as other measures of their experiences while completing the degree program. In addition, it was possible to construct a complete set of academic results for each respondent as well as the full distribution of all grades in the classes they took.

This paper proceeds as follows. First, a survey of the limited literature concerning the *CEQ* and some elements from the extensive literature concerning factors that have been found to influence responses to *Teaching Quality Instrument*s that are pertinent to this analysis is provided. Next, a description of the data that was combined to create a student-level data set for those students who completed the *CEQ* is detailed. This Section consists of five sub-parts: the first details the Good Teaching Scale of the *CEQ* that was modelled and the additional information added from the corresponding *Graduate Destination Survey* (*GDS*) for each student, next the details of the *Teaching Quality Instrument*s in use at The University of Melbourne over the period of the data are provided,

then a description of the student record data available for all the respondents to the *CEQ* is given, followed by a description of the measures of the grade distributions of all the classes taken by the respondent students, and the last data description gives the various ways in which the degree programs completed by the students may be classified. The next major section describes the statistical model used for the analysis and the variables used to predict the responses to the *CEQ*. This is followed by the results section discusses the estimated parameters. Finally, conclusions are presented in the last section.

Literature Review

The literature concerning student surveys is extensive. This review is presented in two parts. The first, takes a cursory look at some of the factors that have been found to influence responses on the *Teaching Quality Instrument*s with particular emphasis on grading policies in order to identify potential factors that may also influence responses to the *CEQ*. The second, examines the limited number of previous quantitative analyses that only use the information collected on the *CEQ*.

A brief overview of the literature concerning *Teaching Quality Instruments.*

Teaching Quality Instruments that are administered in-class (as well as on-line) have been the subject of a wide range of analyses. Many of these have investigated the manner in which they were influenced by various factors that were outside the control of the lecturer and the degree to which these factors could be accounted for to allow for comparisons between lecturers. A consensus derived from the analyses of numerous *Teaching Quality Instrument*s has been their reliability for differentiating class instructors (e.g. Marsh 2007) however the proper interpretation of these surveys requires the consideration of a number of covariates such as those found by Davies et al (2007). Richardson's 2005 survey provided a list of those factors that may be considered confounding or nuisance variables in the determination of average scores for a particular class or lecturer. Although an extensive review of the literature that examines the nature of the responses to traditional in-class

*Teaching Quality Instrument*s lies outside the scope of this analysis, a number of the factors that have previously been investigated are highlighted below.

A major factor that has been found to influence responses to *Teaching Quality Instrument*s are the expectations of the students taking the class. McKeachie (1979, p 390) observed that "Probably the single most important student variable affecting satisfaction is student expectations." Other factors that have been shown to be important in the comparisons of *Teaching Quality Instrument*s are the characteristics of the class taught. These include such measures as: the class size, the year level, if the class is required for further study, the department or discipline in which the class was taught and the distribution of final grades.

A series of studies have examined the relationship between the discipline taught and student evaluations. Feldman (1978) reported that Humanities and Arts-type subjects receive higher ratings than Social Science classes and that these, in turn, received higher ratings than Mathematics and Science classes. Similar results can be found in: Braskamp and Ory (1994); Cashin (1990); Centra (1993); Marsh and Dunkin, (1992); Neumann, (2000); and Sixbury and Cashin, (1995). Although Marsh (1982) concluded that the instructor effect maybe up to five times as large as the effect of the subject matter it was still found to be a significant factor.

Some researchers have been concerned with the existence of a potential for the lecturer-in-charge to buy better evaluations from the *Teaching Quality Instrument*s with lenient grading policies (Clayson, 2009, Matos-Diaz and Ragan, 2010). In Johnson's (2003) treatise concerning the phenomena of *Grade Inflation* he estimated that the grades of students in classes taught in different departments varied by a significant margin. These differences were found by analysing the transcripts of approximately 15,000 students enrolled at Duke University from 1995 to 1999 (page 207).

Achen and Courant (2009) confirmed Johnson's finding and go further by investigating which classes were subject to grade inflation versus grade deflation. They concluded that students who

perceive that they are subject to unfair grading may require significant time commitments from the lecturer. The process by which the lecturer in charge is motivated to acquiesce to pestering students was modelled by Franz (2010). Based on separate surveys of students and of academic staff at the University of California at San Diego, she proposed that the cost of this pestering was greatest for academic staff whose primary evaluation for tenure and promotion was their research output. Thus, the perception that leniency in grading can buy better *Teaching Quality Instrument* results may be due to a rational calculation by the lecturer to minimise the cost of pestering students. In addition, one may assume that the degree of sensitivity by lecturers may vary by department and enrolment size. Achen and Courant (2009) suggest that classes with mathematical results and exact answers such as foreign language structure may be least prone to the perception of unfair grading.

Expectations of grades have also been found to be important for student's perceptions of classes. Nowell (2007) found that students respond to the expected grades of their peers as well as of their own, which would imply that the expected distribution of grades in a particular class may be important. Matos-Diaz and Ragan (2010) discovered that larger expected variations in marks have a negative influence on the results of a *Teaching Quality Instrument* as well. Thus not only the grades received by the student but the expectation of the variation in those grades for others may have influenced their responses on *Teaching Quality Instrument*s.

Literature concerning the *CEQ*

The *CEQ* originated with Ramsden and Entwistle (1981) who designed a survey for the evaluations for courses of study and/or departments within universities that they referred to as the *Course Perceptions Questionnaire*. This initial survey was then refined (see Ramsden 1991a) to become the *CEQ*. A review of history and properties of these various surveys can be found in Richardson (2005). In contrast to the *Teaching Quality Instrument*s, there has been a fairly small literature that examines degree program surveys such as the *CEQ*. A selection of the more prominent studies has been examined below.

Ramsden (2003) proscriptively asserted that the *CEQ* results should not be considered in isolation but in association with other information such as *Teaching Quality Instrument*s. Richardson (2005) compared the *CEQ* to the traditional *Teaching Quality Instrument* and he concluded that due to higher response rates the typical *Teaching Quality Instrument* may provide more valuable information than the *CEQ*. However, one wonders if this observation would be valid in light of the growing use of on-line *Teaching Quality Instrument*s that have been often accompanied by falling response rates as discussed in Dommeyer et al (2004).

Marsh et al (2011) have compared the ability of the *CEQ* to discriminate between department/university combinations to a typical set of *Teaching Quality Instrument* from within a particular institution to distinguish between university lecturers within the same university. They determined that the ability of the aggregate scores of the 2001 *CEQ* to distinguish between approximately 2,150 department/university combinations and 41 universities was fairly limited as compared to the much greater discriminant ability of a *Teaching Quality Instrument* used to compare 195 teachers at a single major US private tertiary institution as reported in Marsh (2007). They also remarked on the potentially lower quality of the *CEQ* data as opposed to the information collected by the typical *Teaching Quality Instrument* due to the poor response rate.

In another line of comparison, Bedggood and Donovan (2012) suggested that surveys such as the *CEQ* and *Teaching Quality Instrument*s appear to confuse measures of student satisfaction with measures of teaching quality. They emphasised that those characteristics of a challenging class may not correspond to the class that leads to the greatest satisfaction at the time when the student is responding to the *Teaching Quality Instrument*. However, after graduation the challenging classes may be viewed with a different perspective and thus surveys such as the *CEQ* could be more likely to measure the quality of instruction than just the level of satisfaction as determined by the *Teaching Quality Instrument*.

Griffin et al (2003) concluded that the *CEQ* was deficient in measuring non-class related elements of the degree program that are also provided by the university. They proposed a modification to the *CEQ* by broadening the scope of the instrument. In particular, these additional scales were designed to capture such dimensions as support for students, the learning community and resources, graduate qualities and intellectual stimulation. The current *University Experience Survey* now used in Australia incorporates these scales in addition to the original scales from the *CEQ*.

As universities increasingly compete for students, the comparisons of the scores from such surveys and rankings have become marketing tools. Since 2012, summaries of the *University Experience Survey* results for all institutions have been available from the Australian government-sponsored *http://myuniversity.gov.au/* website. In a recent paper concerning the marketing of universities, Woodall et al. (2014) concluded that specifically designed marketing oriented surveys may be more appropriate than are surveys such as the *CEQ.*

Although, the *CEQ* has been the subject of a fairly limited set of analyses, none of these studies have attempted to link the aggregate results of earlier *Teaching Quality Instrument*s to the subsequent responses to the *CEQ*. Nor have any of these studies attempted to link the students' academic performance and the characteristics of the classes taken to the responses that they have made as is done in this study.

The Data Used

Three main sources of data from the University of Melbourne were employed in this study: the 2002 to 2006 individual responses to the *CEQ* and the *GDS*, the 1996 to 2005 average values for the *Teaching Quality Instrument* for classes at the University of Melbourne, and student enrolment records for the period from 1995 to 2005 (see Hirschberg et al., 2010, for details).

The student data records used includes both characteristics of the student and the characteristics of the classes they took. First, the responses to the *CEQ* and the *GDS* and their corresponding marks in all the classes they took were combined to create a student record. Then, the

class specific characteristics derived from the average *QOT* scores and the distribution of all student marks in the class were used to create a set of class histories for each of the 52,000 classes taught from 1996 to 2005. Third, the class histories were combined with the student records based on the classes they took to form a total student record that also includes the rank for the student's mark in each of the classes they completed.

The *CEQ* and the *GDS* data

Over the period from 2002 to 2006 10,433 domestic students reported on their experiences in 14,728 separate degree programs in the *CEQ* and the *GDS* at the University of Melbourne. This discrepancy was due to a number of students who completed more than one degree program at the same time.

The *GDS* was designed to collect information from recent graduates about their employment while at university and their subsequent work and study experiences since they completed their degree. It was administered by Graduate Careers Australia at approximately the same time as the *CEQ* however, instead of phone follow-up to a mail survey as was done for the *CEQ* the *GDS* was a solely a mail survey. In this analysis the separate responses for these two surveys were matched to create a single survey record. The components from the *GDS* that were included were: the hours worked, the wage paid, the degree program that the student may have subsequently enrolled in, their employment duties and the industry in which they were employed.

The *CEQ* used in this paper is based on a five-point Likert scale in which respondents are asked to express their degree of agreement on a series of statements concerning the degree(s) that they completed. The scale examined here was the *Good Teaching Scale*, which was defined as the average score on the degree of agreement with the six statements concerning instruction in the student's degree, as listed in Table 1. In addition to computing the average response on these six statements the standard deviation of these six responses was also computed to indicate the degree of consistency of the student's responses to the six statements.

(Table 1 here)

Van Herk et al. (2004) proposed that "response style" could influence variation in responses to questionnaires. A measure of "response style" was defined to indicate the degree to which some individuals were more unequivocal than others in their responses to a Likert scale. For example, in the case of the five-point scales used in the *CEQ*, unequivocal responders were more likely to choose 1 or 5 rather than 2, 3 or 4. To measure this tendency the "extreme response index" was defined as the proportion of all responses in the questionnaire that were either 1 or 5. Thus the index for each respondent was based on their responses on all questions in the *CEQ* and not just the responses in the *Good Teaching Scale* component of the survey.

The Melbourne *Teaching Quality Instrument*

The University of Melbourne (the Quality of Teaching questionnaire or *QOT*) has evolved since its inception in 1995, the core questions have remained largely unchanged. These types of surveys are referred to by many other acronyms in other Australian Universities as detailed in Davies et al. (2010). The equivalent survey is now entitled the *Subject Experience Survey* or *SES*.

Although the wording of the questions in the *QOT* has evolved since its inception in 1995, the core questions have remained largely unchanged. A principal-component analysis of the correlation matrix of the core questions revealed that over 76% of the variation in the responses can be explained by the first component and that the response to question 2 (*Q2* agreement with the statement "*This class was well taught*") has a correlation of 0.92 with this component. In addition, the average response to *Q2* had been employed by the University of Melbourne as the primary evaluation metric of classes and lecturers, thus this average was used as the measure of student evaluation for the classes in the data used here.

The observed average QOT_{ik} score for class *k* taught in semester *i* was decomposed into the level expected for class *k*, $\overline{QOT_k}$ and the difference from the expected level in the semester the student took the class, ΔQOT_{ik}. Thus, the observed QOT_{ik} can be written as the sum:

$QOT_{ik} = \overline{QOT}_k + \Delta QOT_{ik}$. The expected values ($\overline{QOT_k}$) for class *k* were predicted from the regression models based on the specifications used in Davies et al. (2007). They found the key determinants of the responses to *Q2* to be enrolment, year level, the year in which the survey was conducted, and the response rate (the proportion of enrolled students who responded to the survey). These factors were used in a series of auxiliary department specific regressions that were estimated based on the model:

$$QOT_{ik} = \alpha_k + \alpha_i + x_{ik}\alpha_{11} + x_{ik}^2\alpha_{12} + z_{ik}\alpha_{21} + z_{ik}^2\alpha_{22} + \varepsilon_{ik}\,, \qquad (1)$$

where QOT_{ik} are the *Q2*'s average response for class *k* taught in semester *i* and x_{ik} and z_{ik} denote enrolment and the response rate, respectively. Thus, it was assumed that a student taking a class that has been previously taught has an expectation of its average *Q2* response. Using this model the difference between the expected *Q2* and the actual *Q2* for the class can be interpreted as an indicator of the information or surprise at the actual outcome. When the class only appeared once in the data these values are set to zero.

Separate averages of the positive and negative differences were computed over all the classes taken by each student. The average of the positive deviations for all the classes taken by a student was designated as QOT^+ thus $QOT^+ = \frac{1}{n}\sum_{\forall ik} \max(\Delta QOT_{ik}, 0)$ where *n* was the number of classes they took. Conversely, the average of all the negative deviations was defined as QOT^- where $QOT^- = \frac{1}{n}\sum_{\forall ik} \min(\Delta QOT_{ik}, 0)$. These variables were created to provide an insight into how the conduct of a particular class in a particular semester did or did not conform to the student's expectations.

Student Record Data

A set of student records for those students who completed the *CEQ* and the *GDS* containing their marks and the classes they took were provided for this study. Using these values it was possible to construct a set of marks they received and to match the average *QOT* values for all the classes they

took. Since these data were available for the student's entire period of enrolment at Melbourne new variables were also constructed to measure their experiences in the most recent classes they took as well as their averages over their entire time at the University. In some cases, students may have been enrolled in degree programs at the University of Melbourne prior to the degree for which they have completed a *CEQ*. Thus, it was necessary to limit the period over which these characteristics are measured to be able to infer their influence on the degree under examination. Unlike transcripts that record only 13 broad grade classifications such as A+, A, … ,D-, F, the University of Melbourne transcripts include a numeric mark from 100 to 0 as well as one of six grade designations. The six grades are based on the 0 to 100 scale as follows: 49 or lower indicates failure (F), 50–64 is a pass (P); 65–69 is a 3rd class honours (H3); 70–74 is a 2nd class honours division B (H2B); 75–79 is a 2nd class honours division A (H2A); and 80–100 is a 1st class honours (H1). Thus it was possible to measure the proportion of a student's marks that were just below the cut-off values for the next grade in order to establish if the incidence of such close-marks influenced the student's responses to the *CEQ*.

In addition, it was also possible to include the student's rank in each class they took. These were scaled so that 100 would be the top mark and 1 the lowest. By using the distribution of all the marks in their classes from the data described in the next Section, it was possible to determine the student's rank that was adjusted so that ties receive the same rank. The rank can then be used as an alternative to their mark that may be influenced by grade inflation or the class and subject specific grade anomalies observed by Johnson (2003).

The Distribution of Marks by Class

Also made available for this study were the enrolment records by class and the marks for more than 2.1 million student–subject combinations in over 52,000 classes taught at the University of Melbourne from 1996 to 2005. These records included; the class codes, the semester taught, and the

individual marks, but with no student specific identifiers. These records were used to construct the distribution of the marks for each class the survey respondents took.

A significant feature of the grade distributions over this period at The University of Melbourne were the percentage of marks close to the next grade. From the histogram of all the marks in the enrolment sample shown in Figure 1 it can be seen that the drop in the number of marks just below the cut-off points for the next grade at the values of 50, 65, 70, 75 and 80 are quite pronounced. In this analysis these just below marks (47, 48, 49, 64, 69, 74 and 79) are defined as the *close-marks*.

(Figure 1 here)

The drops for these close-marks may have been indicative of class coordinators erring on the side of students or of what some researchers have referred to as "leniency" (Clayson, 2009). One explanation for the lower frequency of the borderline cases maybe the "threat of students' "nuisance", as proposed by Franz (2010). Using the marks in every class it was possible to construct a measure of the prevalence of close-marks in each class as the proportion of all the non-zero marks that are close-marks.

In addition, it was also possible to examine each student's transcript to determine the proportion of all their own grades that were due to close-marks as well. These separate measures facilitated the comparison to be made between the student's reaction to their own experiences and their reaction to the experience of all students taking the same classes as they did.

The Classification of Students by Degree

In order to allow for variation between students when classified by their degree it was necessary to consider a number of ways to classify the degree they received. One method for the classification of degrees was to use the names designated by the institution. However, a major difficulty in classifying students by the names of the degree programs at the University of Melbourne during this time was the large number of degree programs and the wide disparity in their sizes. There

were 363 different degrees reported in the observations used. Table 2 provides the faculty distribution of the number of degrees completed by the students in the sample used here. From Table 2 it can also be noted that the percentages of students completing degrees by faculty in this sample, was fairly constant over these years.

(Table 2 here)

Because most undergraduate students in Arts, Science and Economics & Commerce completed only one of a small number of degrees, the distribution of responding students from these faculties was highly skewed to the extent that approximately 45% of the students were enrolled in only 10 degree programs, and only 1% were enrolled in one of the 95 degree programs with three students or less. For example, over 10% of the students in this sample were classified as having completed a Bachelor of Arts degree even though the majority of their classes were taken in only one department.

(Figure 2 here)

In addition to the degrees defined by the university, a set of five alternative definitions were also used to categorise the degrees received. The six classifications of degrees are listed in Table 3 along with the number of groups defined by each. The associated histograms of the number of students in each group as defined by these classifications are shown in Figure 2. From this figure it can be seen that the classifications defined by departments where the majority of classes were taken (type 5) and the field of education as defined by the Australian Department of Education, Employment and Workplace Relations (*DEEWR*) (type 4) resulted in the most even distribution of students in each grouping variable.

(Table 3 here)

Note that classifications two and three were based on the student's post-graduation experiences as reported in the *GDS* and not on the particular degree they completed. These classifications were defined by the reported duties on job the respondents claimed to have since

finishing their degree and industry in which they are working. A drawback with the use of these categories was the large proportion of responses (almost 25% of the sample) listed as *unknown*.

The classification with the smallest number of sub-groups is type 6 which was defined by the faculty in which they completed their degree and the level of the degree (Bachelor, Bachelor with Honours, or Graduate). These classifications were indicative of the funding by the Australian government that provides university funding on a per capita arrangement with a formula that varies by the level of the degree. The Honours degrees usually involved an additional year of study after a three year undergraduate degree. The graduate classification includes Post-Graduate Diplomas, Master's Degrees and Doctorates.

The Method for Analysis

This section follows Richardson's (2009) methodology to examine the degree to which these experiences influence the observed responses to the *CEQ*. The model estimated belongs to the class of mixed models (Demidenko 2004) and was defined as:

$$\underbrace{y_t}_{\substack{\text{Good}\\ \text{Teaching}\\ \text{Score}}} = \beta_0 + \underbrace{\sum_{i=1}^{K} x_{ti}\beta_i}_{\substack{\text{Student}\\ \text{Records}}} + \underbrace{\sum_{h=1}^{H} w_{th}\pi_h}_{\substack{\text{Class}\\ \text{Histories}}} + \underbrace{\sum_{j=1}^{M} d_{tj}\lambda_j}_{\substack{\text{Degree}\\ \text{Classifications}}} + \underbrace{\sum_{l=1}^{L} z_{tl}\gamma_l + \varepsilon_t}_{\substack{\text{Random}\\ \text{Component}}}, \qquad (2)$$

Where y_t was the *Good Teaching Scale* registered by student *t* based on agreement to the statements listed in Table 1. The w_{th} indicated the *h*th characteristics of the classes taken by student *t* as recorded in the class histories. The x_{ti} indicated the *i*th characteristic of student *t* based on their characteristics and their academic performance. The d_{tj} represented the *j*th fixed effect to account for the degree program of the student *t*. The random component of the model comprises ε_t, was defined as identically and independently normally distributed with a mean of zero, and the random effects γ_l, which were defined as distributed according to the normal distribution with a zero mean. The z_{tl} were defined as quadratic functions of the extreme response index and the standard deviation of the *Good Teaching Scale* computed over the agreement with the six statements in Table 1.

This model was estimated using all the responses from Australian domestic student responses. The sample was limited for two reasons: One, most of the international students returned to their home countries and were thus unreachable and would be severely underrepresented in the sample. Two, the funding decisions of the *Australian Learning Performance Fund* were based solely on the sample of domestic students' *CEQ* responses. However, the class characteristics and *QOT* responses were based on all students' results. The descriptive statistics for the regressors are listed in Table 4.

(Table 4 here)

The x_{ti} were specific to the responding student and included: age, gender, average mark over all classes, the proportion of their marks that were close-marks, the number of classes taken at the same institution, changes in average mark in their last year, their average rank in their classes taken in the last year, their number of hours worked after graduation, their salary after graduation, the year they received the degree, the amount of work they engaged in while studying, the level of further study, the type of qualification earned, and the number of degrees completed simultaneously (up to a maximum of four degrees were reported on with over 45% reporting on two degrees in the same survey).

The w_{th} were those variables that were measures of the characteristics of the classes the responding student took while working on their degrees. They included the average proportion of close-marks earned by all students in the classes taken, the change in the average of *QOT Q2* responses in the classes taken in their last year, the average enrolment in all classes taken, the difference between the average enrolment for the classes taken and the average enrolment in the class from past semesters, and the decomposition of the average *QOT* scores into the average expected, the average of the positive differences from expectations and the average of the negative differences as defined above.

The d_{tj} represented the *j*th indicator variables to account for the particular degree program for the student *t*. As discussed above and summarized in Table 3, there were six different classifications

used to define the completed degree program. To account for these different groupings, six separate models were estimated where each employed one of the degree categorisations to define the degree type fixed effects.

The Results

The estimated coefficients for the model defined in (2) are provided in Table 5. This table reports the coefficient estimates for each of the six models defined by the six alternative degree classifications defined in Table 3. The student specific experience coefficient estimates are reported in rows 1 to 11, the class level coefficient estimates are listed in rows 12 to 18, the *F*-statistics for composite significance tests on the dummy variables (rows 19 to 23) and the random coefficients (rows 24 to 27), as well as the degrees of freedom for each model (row 28) and their associated *adjusted R-squared* ($\overline{R}^2$) values (row 29). The $\overline{R}^2$ values indicate that model 1 was the best-fitting model overall. These $\overline{R}^2$ values indicate that from between 26% to 28% of the variation in the responses was explained by the estimated models. The discussion below provides the implications for the various conditioning variables.

(Table 5 here)

First looking at the variables that were influenced by the student characteristics it can be seen that age, the average rank in the last year, average marks and the change in marks in the last year all were found to be significant factors in the *Good Teaching Scale*. The estimated coefficients on age (in row 1), consistently indicated across all models, that older students evaluated the teaching they received more positively, however from Table 4 it can be seen that 90% of all the ages range from 21 to 30, thus the 9 year difference only results in a difference of at most 4% of one standard deviation of the average *Good Teaching Scale*. This is based on the results for model 3 where the coefficient estimate of .0039 times 9 equals .0316 or approximately 4% of the standard deviation of the average *Good Teaching Scale* of .79. From the estimated coefficients on female (see row 2), gender only appeared to be significant in one model.

Students with higher rankings in the final year (row 3) and in two of the models the higher the average marks (row 4), the higher the value of the *Good Teaching Scale*. However, increases in marks for their last year (row 5) appear to have a negative effect on the *Good Teaching Scale*. In determining the size of these effects it can be shown that 90% of the average ranks in the last year of the degree fell between the 22nd percentile to the 91st percentile which suggests that 20% of the standard deviation in the average *Good Teaching Scale* could be due to the variation in last year's average rank. In addition, 90% of the average marks lay between 59 to 83 which implies a potential variation of 10% of the average *Good Teaching Scale* standard deviation. However, students who had only recently started to receive higher marks in their final year than before were prone to evaluate their degree programs more negatively (row 5). The 90% difference in values for these differences in marks is 18 marks or -8.8% of one standard error of the average *Good Teaching Scale*.

In summary, it appears that students who did well throughout their degree programs were positively disposed to the instruction they received, but those who belatedly achieved higher marks tended to respond more negatively, with ranks having double the influence of marks. Alternative models were estimated with the average of the ranks for the last 3 years and the average marks for the last 3 years in place of these variables and all 6 models resulted in only statistically significant values for the ranks and none for the marks (even at the 10% level).

The number of hours students spent working (row 8) and the log of their post-graduation salary (row 10) appeared to have had little influence on student evaluations. However, students who reported no hours of work at all responded negatively (row 9), except in models two and three where the fixed effects accounted for the employment status.

In examining the Class Level factors that were outside the control of the student that measured the characteristics of the classes they took it can be noted that the average *QOT*, the negative deviations from the average *QOT*, the proportion of the marks in the class that are close to

the next level and the average enrolments in the classes all were found to have a significant impact on the *Good Teaching Scale*.

The average in-class survey results for the classes taken by the respondent ($\overline{QOT}$ in row 12) had a positive impact on the *CEQ* in those models that do not explicitly account for the major field of study or department, in models 4 and 5. 90% of the average *QOT*s range from 3.6 to 4.3 thus the impact would be 23% of a standard deviation of the average *Good Teaching Scale*. It was found that even after accounting for the degree program taken, $\overline{QOT_k}$ could have a significantly positive impact on the *CEQ* responses. This may have been due a wide variation within degree programs of the expectations for the *Q2* that were not accounted for. Across all models, the effect of the *QOT* below expectations (QOT^- in row 13) was unambiguously negative since 90% of the values for this variable were recorded between 0 and -.32 this implies that a drop of up to 23% of a standard deviation in the average *Good Teaching Scale* could be due to the lower than expected *QOT* in the subject taken. Note that the positive deviations from the expected value of the *QOT* (QOT^+ in row 14) did not have an estimated coefficient that was significantly different from zero in any model. Hence, this result appears to have uncovered an asymmetry in this relationship between the in-class *QOT* ratings and the overall degree ratings from the *CEQ*: higher than expected *QOT* ratings do not result in higher *CEQ* values for the *Good Teaching Scale*, but lower than expected *QOT*s have a significant negative impact on the *CEQ* that can counteract the positive impact of the average *QOT*.

When comparing the student level factors to the class level factors an interesting dichotomy appeared. From row 6 it was found that the proportion of a respondent's own close-marks were not a significant factor in explaining the *Good Teaching Scale* responses. However, the average proportion of all the marks in the classes taken by the student that were close-marks (row 16) was a significant negative factor in all models. This result appears to be related to the findings by Nowell (2007) and Matos-Diaz and Ragan (2010) that other student's grades were important factors in class evaluations and that this variable may provide an indication of "leniency" in the class grading. In this case, 90%

of the range of proportions of close-marks was from 0 to .28 and could account for -23% of the standard deviation of the average *Good Teaching Scale*, an effect of the same magnitude as the *QOT*.

A consistent finding across the models was that average class enrolment (row 18) had a negative impact on the *Good Teaching Scale*. In this case the 90% range of 21 to 429 indicated a –33% difference in the standard deviation of the average *Good Teaching Scale*. This result was consistent with analyses of factors that influence subject-specific *QOT*s (see, e.g., Davies et al., 2007) which find that once controlling for other factors larger class sizes have a negative impact on evaluations.

An innovation of this study was the inclusion of response style and a measure of consistency in the responses that make up the dependent variable as random factors in the model. It was found that the random component was positively influenced by the extreme response-style measure and negatively by the measure of the internal consistency of the responses on the *Good Teaching Scale* in a nonlinear way. The more extreme the responses on all the questions in the *CEQ* the higher the *Good Teaching Scale* response, at the average value one would find approximately 25% of the standard deviation of the average *Good Teaching Scale* due to the extreme response index. The standard deviation of the responses to the agreement or disagreement with the six statements listed in Table 1 were used as a measure of the consistency of student's responses. It was found that the greater the inconsistency of the *Good Teaching Scale*, the lower the value predicted *Good Teaching Scale.* In this case the average of the standard deviation of the six values would have approximately a -30% of the standard deviation of the dependent variable effect. To account for the use of a quadratic for these variables the marginal influence varies and they are evaluated at the average for these variables.

In addition to the coefficients on the student specific continuous variables, a series of related indicator variables were also tested using a composite test that all the categories for: the year of the survey (row 19), the combinations of part-time and full-time study and different levels of outside

work (row 20), the level of the further qualification they were now enrolled in (if any) (row 21), the level of the degree they completed (row 22) and the number of degrees they were reporting on in this survey (row 23). Except for the year of the survey all the other factors were significantly different from each other at the 1% level in all models but one where we condition on the level of the qualification (model #6). Table 5 lists the F-statistics for each case along with the numerator degrees of freedom. The denominator degrees of freedom for these statistics appear in row 28.

In order to interpret the parameter estimates for the indicator variables tests were conducted to determine if differences between the predicted *Good Teaching Scale* for each category were significantly different from each other at the 5% level. From this analysis it was found that the largest difference between those respondents that studied full-time and were not working and those that studied part-time and were not working was .15 or almost 19% of one standard deviation of the average *Good Teaching Scale*.

The same method was used to compare post-graduation degree programs. It was found that those respondents who were enrolled in Master's Degrees and Doctorates on average scored the *Good Teaching Scale* higher than students those who were enrolled in Certificate programs. The model with the greatest difference resulted in an average of 27% of one standard deviation of the average *Good Teaching Scale* higher score for a Master's Degree program over a Certificate program. In comparing the coefficients for the degree that they just completed it was found that the predicted *Good Teaching Scale* differential was the greatest for those that completed a non-honours Bachelor degree as opposed to those that completed an Advanced Diploma. In this case the difference was 30% of the standard deviation of the average *Good Teaching Scale*.

Conclusions

In this paper, the responses to the *Good Teaching Scale* of the degree program survey entitled the *Course Evaluation Questionnaire* (*CEQ*) were matched to the experiences of the graduates in the classes they took in order to predict how class characteristics influenced their assessment. A number

of factors were found to influence the *Good Teaching Scale*. These include measures of a student's performance, their subsequent employment or further study, the measures of teaching quality used and other characteristics of the classes taken.

A major implication of the findings was that a subject's average score on the primary question ("Was the class well taught?") of the local *Teaching Quality Instrument* appeared to influence responses to the *CEQ*. However, this effect was only apparent if the fixed effects did not control for the departments in which students took most of their classes (as reported in models 4 and 5). This mirrors the widely reported observation that average *Teaching Quality Instrument* scores are strongly influenced by the department in which a class is taught (See Feldman 1978, Braskamp and Ory, 1994; Cashin, 1990; Centra, 1993; Marsh and Dunkin, 1992; Neumann, 2000; Sixbury and Cashin, 1995). Thus, one can conclude that students who took classes within a particular department form expectations about the delivery of those classes.

There was also evidence of an asymmetric influence when class evaluations for a particular semester when a class was taught deviated from student expectations. Although one might expect that positive deviations from the average in-class *Teaching Quality Instrument* would result in higher *CEQ* from these students this was not found. However, if the in-class *Teaching Quality Instrument* scores were lower than the average expected scores for the class, the students reacted with a significant negative reaction. This result implies that the policy should be focused on maintaining a consistent *Teaching Quality Instrument* that avoids major dips in performance more than concerns about very high results.

The sizes of the classes along with measures of the distribution of marks in the classes taken were also used to characterise the in-class experiences of the respondents to the *CEQ*. In particular, the distribution of marks was used to define the degree of lenience or lack of it based on the proportions of close-marks just below the cut-off mark required for the next grade. The estimated coefficients for these measures indicated that students have a negative reaction to taking classes in

which they do not get the "benefit of the doubt" or are not treated leniently. However, the aversion to close-marks was based on the measure of the class averages of these marks rather than students' own experiences.

In addition to the influence of the way in which the classes were taught as measured by the *Teaching Quality Instrument* and the distribution of grades a number of other factors were found to influence the *CEQ* results. It was found that not only a student's marks but also their rank within a class can influence their responses to the evaluation survey. In particular, a student's average rank in classes they took in their final year of study was found to have a significant impact on their perceptions of their course of study. While it was also found that higher marks do not have an unambiguously positive impact on student responses. Although students' average marks positively influence their evaluations, students whose marks changed in their final year of study were less prone to evaluate teaching positively.

By matching the responses to the *GDS* and the *CEQ* it was possible to determine whether student responses were affected by their full-time/part-time status and/or by whether they went on to study for a higher degree. Students who studied full time in their final year and those who subsequently studied for a higher degree were found to give more positive responses than did part-time students and those who did not go on to further study. This finding contrasts with those of Duarte et al. (2012), who found that the opinions of former students do not markedly differ from those of current students.

An additional aspect of this analysis, which has typically not been used in the evaluation of surveys such as the *CEQ*, was the modelling of random effects to account for the response style of the student and the internal consistency in the *Good Teaching Scale* based on the variance of the responses to the underlying questions. Both of these factors were found to influence the errors in the model specifications considered here and indicate that they should be included in future analysis of similar data.

Just as previous research has found that end-of-semester *Teaching Quality Instrument*s were influenced by student performance and other non-teaching aspects of a subject this study has found that a degree program survey was also subject to factors that were not directly related to the program's delivery or the classes taken. The $\overline{R}^2$ values reported in Table 5 indicate that up to 28% of the variation in the *CEQ* could be explained by a student's experience and characteristics, which implies that one can adjust these averages to account for the influences of these factors by the use of this type of model. Although evidence that class surveys can be used to predict degree surveys was uncovered, the links were neither direct nor symmetric. In addition, a descriptive analysis of the *CEQ* return rates by subject found that students with negative experiences might be overrepresented in the sample as detailed in Section 5.2.3 of Hirschberg et al (2010). Thus the interpretation of post-graduation surveys such as the *CEQ* without consideration of the appropriate conditioning variables should be discouraged. Consequently, policy decisions should not be made on simple average cross-institutional or inter-institutional comparisons of the results of post-graduation surveys without taking the factors examined in this study into account.

Acknowledgement

Partial support for this research was provided by grant CG7-384 from the Australian Learning and Teaching Council Ltd, an initiative of the Australian Government and a grant from the Teaching and Learning Performance Fund of the University of Melbourne. We particularly acknowledge 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 and Dr. David Morton for editorial comments. We also wish to thank attendees at the following conferences where earlier versions of this paper were presented. The World Universities Forum 2011, Hong Kong, January 14th, 2011, The 8th 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 29th of June 2010. We also wish to thank the editor and the anonymous referees for their comments that have improved the presentation of this work. The views expressed in this report do not necessarily reflect the views of the University of Melbourne, the Australian Learning and Teaching Council or the Australian Government.

References

Achen, A. C., and P. N. Courant, (2009), "What Are Grades Made Of?", *Journal of Economic Perspectives*, 23, 77-92.

Bedggood, R. E. and Donovan, J. D. (2012), "University Performance Evaluations: What are we Really Measuring?", *Studies in Higher Education,* 37, 825–842.

Braskamp, L. A., and Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.

Cashin, W. E. (1990), "Students do rate different academic fields differently", In M. Theall and J. Franklin (Eds.), *Students ratings of instruction: Issues for improving practice: New directions for teaching and learning* (113-121). San Francisco: Jossey-Bass.

Centra, J. A. (1993), *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*, San Francisco: Jossey-Bass.

Clayson, D. (2009), "Student Evaluations of Teaching: Are They Related to What Students Learn? A Meta-Analysis and Review of the Literature", *Journal of Marketing Education*, 31, 16–30.

Davies, M., Hirschberg, J., Johnston, C. and Lye, J. (2007), "Systemic Influences on Teaching Evaluations: The Case for Caution", *Australian Economic Papers*, 46(1), 18–38.

Davies, M., Hirschberg, J., Lye, J. and Johnston, C. (2010), "A Systematic Analysis of Quality of Teaching Surveys", *Assessment and Evaluation in Higher Education*, Vol. 35, No. 1, 87–100.

Demidenko, E., (2004), *Mixed Models: Theory and Applications*, John Wiley & Sons.

Dommeyer, C. J., Baum, P., Hanna, R. W., and Chapman, K. S., (2004), "Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations", *Assessment and Evaluation in Higher Education*, 29, 612-623.

Duarte, P. O., Raposo, M. B. and Alves H. B. (2012), "Using a Satisfaction Index to Compare Students' Satisfaction During and After Higher Education Service", *Tertiary Education and Management, 18, 17–40.*

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.

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.

Franz, W. I., (2010), "Grade inflation under the threat of students' nuisance: Theory and evidence", *Economics of Education Review*, 29, 411-422.

Graduate Careers Australia (2009), Graduate Course Experience 2009, *The Report of the Course Experience Questionnaire*, Melbourne, Australia: Graduate Careers Australia Ltd.

Griffin, P., Coates, H., McInnis, C. and James, R. (2003), "The Development of an Extended Course Experience Questionnaire", *Quality in Higher Education*, 9, 259–266.

Hirschberg, J., Lye, J., Davies, M. and Johnston, C. (2010), *Measuring Student Experience: Relationships between Teaching Quality Instruments and Course Experience Questionnaire (CEQ), Australian Learning and Teaching Council*, *http://www.olt.gov.au/project-measuring-student-experience-melbourne-2007*.

Johnson, V. E., (2003), *Grade Inflation: A Crisis in College Education*, Springer, New York, NY.

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. (2007),"Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years", *Journal of Educational Psychology*, 99, 775-790.

Marsh, H. W., and Dunkin, M. (1992), "Students' evaluations of university teaching: A multi-dimensional perspective", In J. C. Smart (Ed.), *Higher education: Handbook on theory and research,* 143-234. New York: Agathon Press.

Marsh, H. W., P. Ginns, A. J. S. Morin, B. Nagengast, and A. J. Martin, (2011), "Use of Student Ratings to Benchmark Universities: Multilevel Modeling of Responses to the Australian Course Experience Questionnaire (*CEQ*)", *Journal of Educational Psychology*, 103(3), 733-748.

Matos-Diaz, H. and J. F. Ragan, (2010), "Do student evaluations of teaching depend on the distribution of expected grade?", *Education Economics*, 18, 317-330.

McKeachie, W. J., (1979), "Student Ratings of Faculty: A Reprise", *Academe*, 65, 384-397.

Neumann, R. (2000), "Communicating student evaluation of teaching results: Rating interpretation guides", *Assessment and Evaluation in Higher Education,* 25(2), 121-134.

Nowell, C., (2007), "The Impact of Relative Grade Expectations on Student Evaluation of Teaching", *International Review of Economic Education*, 6, 42-56.

Radloff, A., Coates, H., Taylor, R., James, R., and Krause, K., (2012), *University Experience Survey* (*UES) National Report*, ACER, Melbourne, http://www.acer.edu.au/documents/UES_2012_National_Report.pdf

Ramsden, P. and Entwistle, N. (1981), "Effects of Academic Departments on Students' Approaches to Studying", *British Journal of Educational Psychology*, 51, 368–383.

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), "Student Surveys and Quality Assurance", *Proceedings of the Australian Universities Quality Forum 2003*, Melbourne, Victoria, 126–135.

Richardson, J. (1994), "A British Evaluation of the Course Experience Questionnaire", *Studies in Higher Education*, 19(1), 59–68.

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

Richardson, J. (2009), "What Can Students' Perceptions of Academic Quality Tell Us? Research using the Course Experience Questionnaire", in M. Tight, *The Routledge International Handbook of Higher Education*, 199–209, Hoboken: Taylor and Francis.

Sixbury, G. R., and Cashin, W. E. (1995), "Description of database for the idea diagnostic form", In *Idea paper, vol. 9*: Center for Faculty Evaluation and Development, Manhattan, Kansas: Kansas State University.

The Good Universities Guide, Hobsons publishing (*www.gooduniguide.com.au*), Melbourne, Australia.

Trigwell, K. and 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, H., Poortigna, Y. and Verhallen, T. (2004), “Response Styles in Rating Sales: Evidence of Method Bias in Data From Six EU Countries”, *Journal of Cross-Cultural Psychology*, 35, 346–360.

Woodall, T., Hiller, A. and S. Resnick, (2014), “Making sense of higher education: students as consumers and the value of the University Experience”, *Studies in Higher Education*, 39, 48-67.

Figure 1: The distribution of all non-zero marks issued to students in this sample.

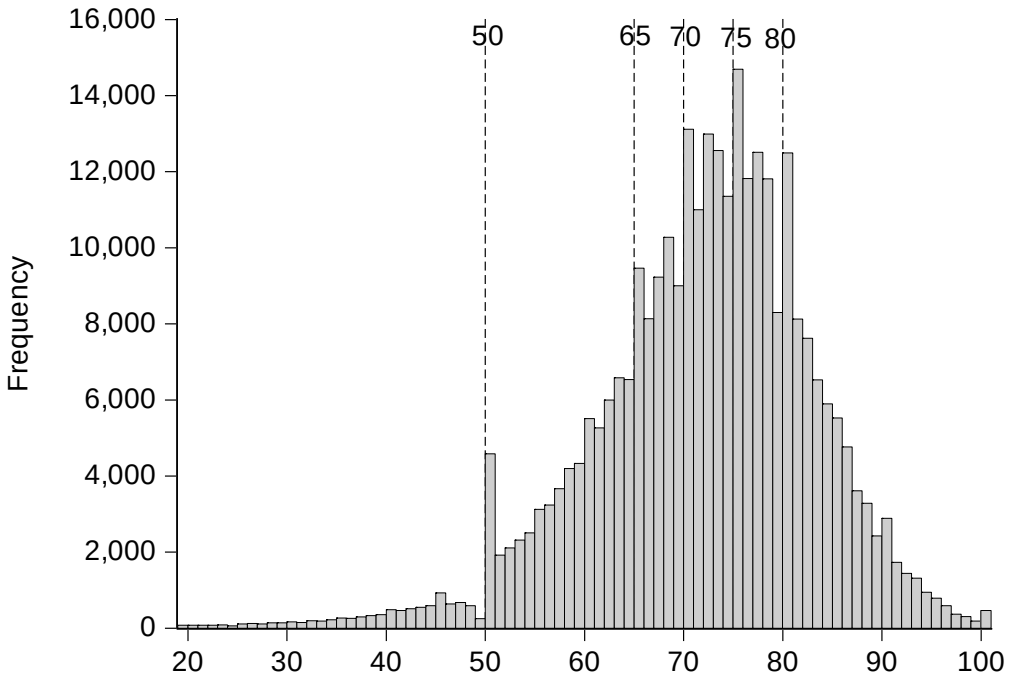
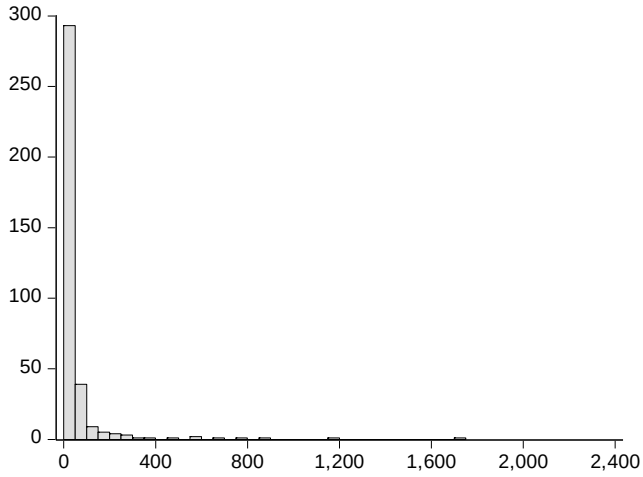
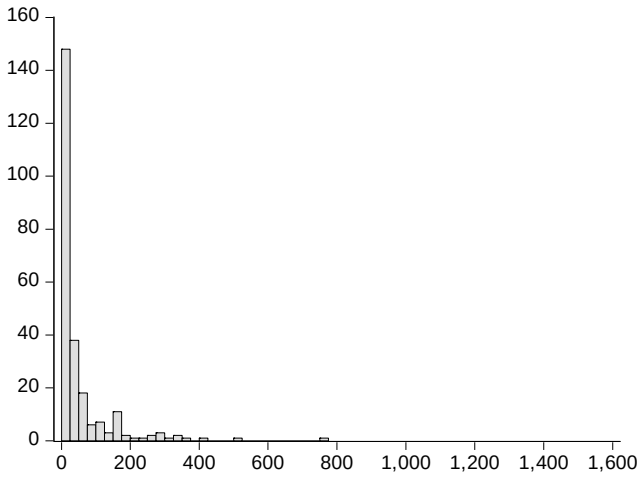


Figure 2: The histograms of the number of categories by size as defined by the number of students for the six fixed effects (note that duties and industry exclude the unknown category)

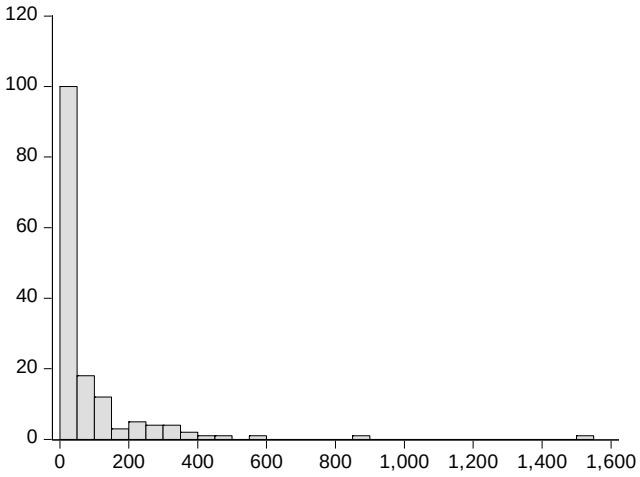
1. Courses as defined by University of Melbourne



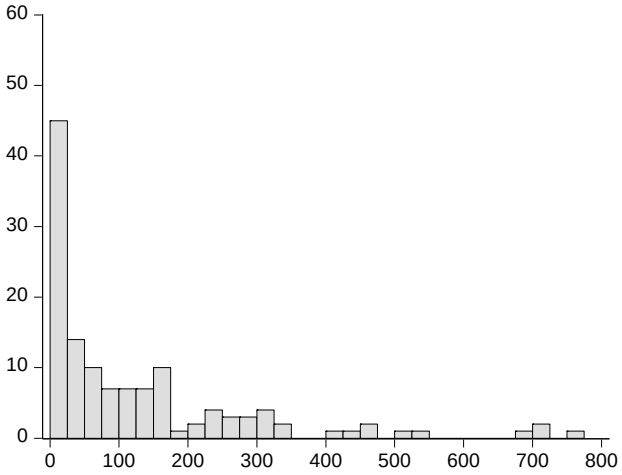
2. Duties on job



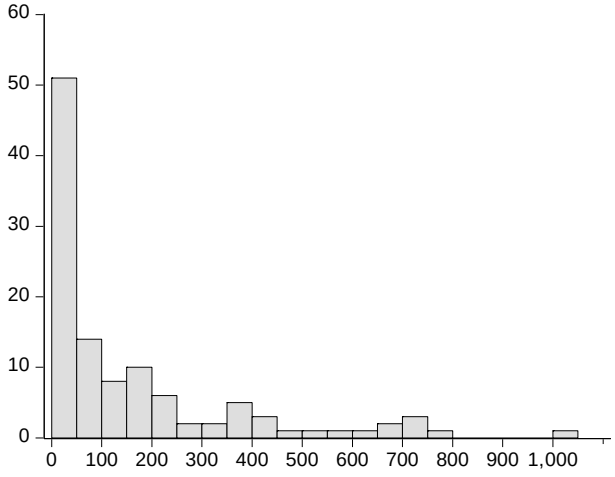
3. Industry



4. Field of education



5. Two departments in which most classes were taken



6. Faculty of course and level of degree

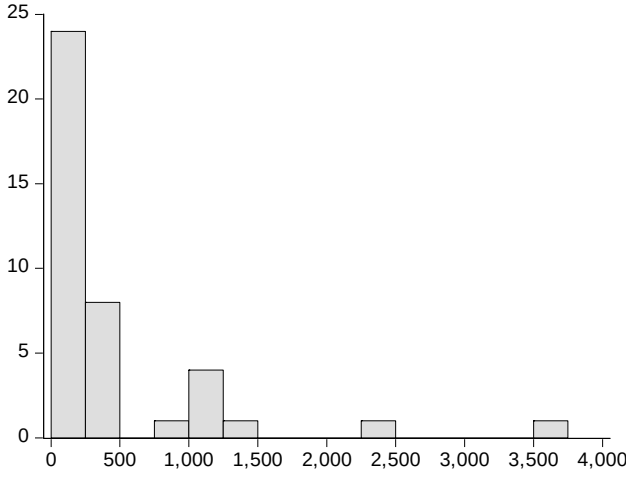


Table 1 The *Good Teaching Scale* is defined as the average agreement with the following statements.

	Statements concerning teaching in the *CEQ*
1	The staff put a lot of time into commenting on my work.
2	The teaching staff normally gave me helpful feedback on how I was going.
3	The teaching staff of this course motivated me to do my best work.
4	My lectures were extremely good at explaining things.
5	The teaching staff worked hard to make their classes interesting.
6	The staff made a real effort to understand difficulties I might be having with my work.

Table 2 The distribution of *CEQ* respondents by Faculty.

Faculty	***2003***	***2004***	***2005***	***2006***
Agriculture	4.48	4.21	2.89	3.72
Architecture	4.65	4.64	4.05	4.02
Arts	19.65	23.04	22.68	19.96
Economics & Commerce	11.10	10.70	11.91	11.96
Education	5.75	5.62	4.16	4.16
Engineering	8.21	7.47	7.16	8.52
Law	5.35	4.83	4.83	4.96
Medicine	5.72	5.93	6.27	7.42
Music	1.71	2.07	2.21	2.08
Science	29.94	29.16	30.70	30.53
Veterinary	1.01	1.04	1.03	0.89
Vic College of The Arts	2.43	1.29	2.11	1.78
Total	100.00	100.00	100.00	100.00

Table 3 The alternative classification of degrees

Classification	***Description***	***Number Groups***
1	Degrees as defined by University of Melbourne	386
2	Duties at job since course completion	292
3	Industry in which graduate currently employed	200
4	*DEEWR* defined field of education	172
5	The two departments in which majority of classes taken	226
6	Faculty of course and level of degree	79

Table 4 Summary statistics for data used in regression modelling.

Variable	*Mean*	*Std Dev*	*Min*	*5%*	*95%*	*Max*
Good Teaching Scale	3.37	0.79	1	2	4.3	5
Number of classes taken	23.29	11.76	1	4	43	71
Age at time of survey	23.85	4.52	19	21	30	76
Flag for female student	0.60	0.49	0	0	1	1
Number of degrees	1.60	0.69	1	1	3	4
Avg Mark over all classes	72.08	7.47	7.00	59.24	83.69	97.88
Avg rank in last year (100 for top)	58.22	20.77	0.56	22.29	90.88	100.00
Change in avg mark in last year	1.72	6.04	-53.00	-6.67	11.00	74.00
Prop of close marks received by student	.13	.07	0.00	.03	.25	100.00
Working hours	25.74	19.20	0.00	0.00	50.00	99.00
Log (starting salary +1)	6.74	4.91	0.00	0.00	10.87	13.38
Zero Hours reported	0.24	0.42	0	0	1	1
Zero Salary reported	0.34	0.47	0	0	1	1
Expected QOT Q2 for classes $(\overline{QOT})$	3.94	0.23	2.83	3.60	4.34	4.82
Negative diff from QOT Q2 (QOT^-)	-0.10	0.11	-1.25	-0.32	0.00	0.00
Positive diff from QOT Q2 (QOT^+)	0.02	0.07	0.00	0.00	0.14	0.73
Enrolment - average enrolment	26.44	69.04	-626.37	-16.62	130.09	2337.36
Average Enrolment	168.11	130.27	3.25	21.14	428.79	987.67
Change in avg QOT in last year	0.06	0.26	-1.86	-.36	.49	1.69
Prop of Close marks in classes taken	.12	.10	0.00	0.00	.29	100.00
Avg Extreme response style Factor (all questions)	0.23	0.20	0.00	0.00	0.87	1.00
Standard Deviation of Good Teaching Scale	0.25	0.13	0.00	0.00	.48	1.05
% of subjects in Dept 1	56.86	24.69	9.52	25.00	100.00	100.00
% of subjects in Dept 2	19.12	11.93	0.00	0.00	38.46	50.00

Table 5 The parameter estimates and F-tests for the mixed models for the *Good Teaching Scale* for Australian citizens and permanent residents

	Effect	*Model Specification for Good Teaching Scale* [a]					
	Label	**1**	**2**	**3**	**4**	**5**	**6**
	Student Level						
1	*Age at time of survey*	0.0022	0.0034**	0.0039***	0.0033**	0.0032**	0.0032**
2	*Flag for female student*	–0.0048	0.0011	0.0023	–0.0277**	–0.0211	–0.0068
3	*Avg rank in last year*	0.0023***	0.0020***	0.0021***	0.0019***	0.0021***	0.0018***
4	*Avg Mark over all classes*	0.0018	0.0023	0.0020	0.0029**	0.0023	0.0032**
5	*Change in avg mark last year*	–0.0039***	–0.0035***	–0.0034***	–0.0035***	–0.0038***	–0.0032***
6	*Prop of close marks received*	0.0149	0.0326	0.0340	0.0178	0.0011	0.0158
7	*If zero hours reported*	–0.0575**	0.0354	0.0396	–0.0513*	–0.0551*	–0.0487*
8	*Working hours*	–0.0006	–0.0001	0.0002	–0.0005	–0.0005	–0.0003
9	*If zero salary reported*	–0.0984	–0.0752	–0.0689	–0.1055	–0.1522	–0.1559
10	*Log of salary + 1*	–0.0097	–0.0086	–0.0068	–0.0098	–0.0144	–0.0147
11	*Number of classes taken*	0.0240	0.5350	0.4350	0.6330	1.0900	0.8410
	Class Level						
12	*Expected QOT Q2 for classes* $(\overline{QOT})$	0.1661***	0.2339***	0.2694***	0.0574	0.0454	0.3077***
13	*Negative diff from QOT Q2* (QOT^-)	0.3800***	0.5531***	0.5099***	0.3942***	0.4791***	0.4508***
14	*Positive diff from QOT Q2* (QOT^+)	0.0034	–0.0115	0.0387	0.0359	0.0803	0.0786
15	*Change in avg QOT in last year*	0.0153	–0.0016	0.0205	0.0151	–0.0062	0.0187
16	*Prop of close marks in classes taken*	–0.1424*	–0.1526**	–0.1499**	–0.1711**	–0.1822**	–0.1565**
17	*Enrolment – average enrolment / 1000*	0.1630	–0.0100	–0.0008	0.1680	0.0760	0.0510
18	*Average enrolment / 1000*	–0.1900***	–0.2300***	–0.2600***	–0.1500**	–0.1500*	–0.1400**
	F-statistics (df for numerator)						
19	*Year survey conducted (df=3)*	0.9604	1.2334	1.4582	0.4778	0.3764	0.6034
20	*Work and Study (df=5)*	4.1533***	4.1684***	4.6517***	3.8362***	4.0302***	3.9466***
21	*Level of further study (df=10)*	2.8683***	2.9481***	2.5587***	3.3495***	2.9387***	2.2211**
22	*Level of qualification (df=4)*	5.0076***	7.6190***	13.1019***	10.5438***	6.3395***	1.5400
23	*Number of degrees (df=3)*	10.2148***	9.4293***	8.8713***	7.1662***	7.0273***	6.4857***
	Error terms						
24	*Extreme response index*	0.3951***	0.4868***	0.5124***	0.4923***	0.4971***	0.5132***
25	*Extreme response index*^{2}	1.2906***	1.1857***	1.1668***	1.1440***	1.1363***	1.1459***
26	*St dev of Good Teaching Scale*	–0.7958***	–0.7622***	–0.7885***	–0.7622***	–0.7581***	–0.7748***
27	*St dev of Good Teaching Scale*^{2}	–1.0326***	–1.0845***	–1.0887***	–1.0952***	–1.0890***	–1.0855***
28	***Degrees of Freedom*** [b]	12,588	12,682	12,774	12,802	12,748	12,895
29	***Adjusted R-squared*** ($\bar{R}^2$)	0.2821	0.2675	0.2647	0.2703	0.2713	0.2628

*a. Significance: *0.1 < prob(t) < 0.05, **0.05 < prob(t) < 0.01, and *** prob(t) < 0.001.*

b. Total number of observations 12,974.