THE UNIVERSITY OF MELBOURNE

MELBOURNE

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 953

NOVEMBER 2005

**SYSTEMATIC INFLUENCES ON TEACHING
EVALUATIONS: THE CASE FOR CAUTION**

by

Martin Davies, Joe Hirschberg, Jenny Lye,
Carol Johnston & Ian McDonald

Department of Economics
The University of Melbourne
Melbourne   Victoria   3010
Australia.

Systematic Influences on Teaching Evaluations: The Case for Caution

Martin Davies, Joe Hirschberg, Jenny Lye, Carol Johnston, and Ian McDonald[1]

Teaching and Learning Unit Faculty of Economics and Commerce, Department of Economics, Department of Economics, Teaching and Learning Unit Faculty of Economics and Commerce, Department of Economics, University of Melbourne, respectively.

Abstract

The evaluation of teaching and learning has become an important activity in tertiary education institutions. Student surveys provide information about student perceptions and judgments of a particular subject. However, as is widely recognised, the appropriate interpretation of this data is problematic. There is a large literature, mainly for the US, on the use and usefulness of student subject evaluations. This literature has highlighted a number of 'mitigating factors' such as subject difficulty, discipline area, etc., that should be taken into account in interpreting the results of these questionnaires. In this paper we examine 8 years of QOT responses from an Economics Department in an Australian University which accounted for more than 79,000 student subject enrolments in 565 subjects. The purpose of this analysis is to establish how the information contained in these data can be used to interpret the responses. In particular, we determine to what extent other factors besides the instructor in charge of the subject have an impact on the raw average student evaluation scores. We find that the following characteristics of the students in these classes had an influence on the average QOT score: year level, enrolment size, the quantitative nature of the subject, the country of origin of the students, the proportion that are female, Honours status of the student, the differential in their mark from previous marks, quality of workbook, quality of textbook and the relative QOT score versus other subjects taught at the same time. However, a number of other factors proposed in the literature to be important influences were found not to be. These include the student's fee paying status, whether they attended a public, private or catholic secondary school, which other faculty within the University they came from, and if the subject was taught in multiple sessions.

1.   Introduction

It is now widespread practice in tertiary institutions to use student feedback in the form of student evaluation surveys to evaluate teaching. The purpose of obtaining such feedback is threefold (see eg. Kember et al 2002): 1) to improve the quality of teaching, 2) to be used in appraisal exercises including tenure and promotion decisions and 3) often it is an explicit requirement by university administrators.

The use of these surveys has often been criticised. It has been argued that students are not the best source to obtain data. Students may have different perceptions of what is important in teaching as they do not have the knowledge necessary for its evaluation and consider it a chore (Simpson and Siguaw, 2000). They are not necessarily best judges of their instructors' performance (Casey, Gentile and Bigger, 1997) viewing it "from very limited or even tainted perspectives" (p. 472). In addition there is now also a large published literature that has examined how student evaluations of teaching can be affected by various influences. This includes teacher or instructor-level determinants, student-level determinants and subject-level determinants[2].

Previous studies have shown that instructor characteristics influence the outcomes of student evaluations. Marsh (1987) concludes that student evaluations are primarily a function of the instructor who teaches a subject rather than the subject that is being taught. Aigner and Thum (1986) conclude from their analysis that 67% of the variation in the student's evaluations of economics instructors' performance was explained by instructor-specific variables including enthusiasm, ability to generate interest in students and interaction between instructor and students. A wide ranging number of studies show that consistently the most important characteristics associated with the overall rating are "*organization*" and "*communication skills*"

---

[2] For a more detailed review of the literature and an overview of our econometric analysis see Davies et al (2005).

(Nelson and Lynch 1984; De Canio 1986; Boex 2000; Gokcekus 2000). In particular, students appreciate instructors who are knowledgeable, warm, outgoing and enthusiastic (Murray, Rushton and Paunonen 1990). It is also thought that these same traits are likely to make the person a more effective teacher, so that students are stimulated to greater achievement and learning. If students feel they have learned in the subject they will give higher ratings. Entertaining teachers do not necessarily receive higher overall student ratings (Costin, Greenough and Menges 1971; Marsh and Ware 1982; McKeachie 1978). However, instructor charisma was found to be a factor that influences student's ratings (Shevlin et al 2000). Bosshardt and Watts (2001) found that economics students gave higher ratings on all evaluated items of teaching effectiveness to those instructors that spoke English as a first language than to those for whom English was not their native language. Some studies show that adhering to a gender appropriate teacher style may be rewarded by higher ratings (Basow and Silberg 1987; Marsh and Dunkin, 1992). Other instructor characteristics have little influence on the overall rating such as for example, the title, degree and position of instructor (Gokcekus 2000) and the years of teaching experience (Jacobs 2002).

Notwithstanding the influence of the instructor, student-specific and subject characteristics could also influence the overall ratings of instructors and are not in general under the direct influence of the instructor. Worthington (2002) in his analysis of evaluation rankings of one instructor in a finance subject found that the age of the student did have some influence on the ratings given. Worthington (2002) also found that higher ratings were likely to be given by students who were expecting a higher grade and were from a non-English speaking background. Aigner and Thum (1986) found that students expecting high grades in a subject tended to give higher ratings than did students expecting lower grades. Boex (2000) found a positive correlation between student grades and instructor ratings with similar conclusions made by Greenwald and Gillmore (1997a), Mason et al (1995) and Nelson and Lynch (1984). A

number of studies also found workload and the difficulty of subject to be positively correlated with instructor ratings. However, the evidence suggests that student evaluations are not unduly influenced by student's personal characteristics such as gender or academic ability (Abrami, Leventhal and Perry 1982; Cuseo 2002; Seldin 1993).

Ting(2000) and Williams and Ory (1992) found that class size is not a very important factor affecting the validity of student ratings. However, an Australian study conducted across four disciplines found that larger classes were rated lower (Neumann 2000). A 20 year study of education subjects found that changes to subjects over time including larger class size and the use of sessional staff were followed by lower evaluation scores from students. Feldman (1978) found that Humanities and Arts-type subjects receive higher ratings than Social Science type subjects although these subjects receive higher ratings than Mathematics-type subjects. Elective subjects were found to be associated with higher ratings than compulsory subjects (Gokcekus 2000; Feldman 1978; Braskamp and Ory 1994). Liaw and Goh (2002) found evaluations not to be related to the time of day or part of the year when the subject was conducted. Furthermore, Feldman (1978) found that evaluations were not greatly affected by time during the term when they are collected.

Typically the results of student evaluations are reported using simple summary statistics including the raw average score. It is common practice for the raw average of the question relating to teaching effectiveness to be compared across instructors and in many instances to a particular benchmark value. Such a practice implicitly assumes that any potential mitigating factors (such as e.g. the effect of student or subject characteristics) have no impact on student evaluation differentials, or that such differentials cancel out in all cases.

In this paper we construct a data set for the period from 1995 to 2003 using 3 sources of data including: the average evaluation scores as reported for all subjects taught in the Economics Department; student records for all students who have taken an economics subject and additional

subject specific information as to the nature of the subject, its class size, time taught and other such information. Our aim is to establish how those factors which are not under the control of the instructor have an impact on the average score attributed to their teaching efforts. We develop a statistical methodology whereby we can condition these average scores in order to make comparisons across instructors. The rest of the paper proceeds as follows. In Section 2 the construction of the data set is discussed and some interesting features of the data are highlighted. The model specification and estimates are presented in Section 3. Section 4 presents a method for evaluating Instructor performance and in Section 5 conclusions are presented.

2.   Data

In each semester since 1994 and in all subjects taught at the University of Melbourne a Quality of Teaching (QOT) survey is undertaken to record students' opinion on the quality of teaching in their subjects. This survey consists of 9 core questions. Additional questions may be added by departments or faculties. Summary results of the core questions are reported to departments and faculties. The Department of Economics is one of the largest departments in the University and includes the area of economics, econometrics and actuarial studies.[3]

The Department of Economics teaches first year core subjects in economics and quantitative methods which are compulsory for all students enrolled for the BCom and BCom (Management) degrees. A number of core economics and quantitative subjects and some elective subjects are taught in second year. At the third year level the department provides over 25 optional subjects of which some may be prerequisites for higher year level subjects. In addition there are a number of fourth (honours) year subjects, a Masters by course work and a three year PhD programme with one year of subject work. The student population is large and diverse. For example, in 2003 the total enrolment in subjects taught by the Department of Economics was 7,984 of which 28% were full fee-paying international students (Annual Report

---

[3] In the analysis reported in this paper only economics and econometrics subjects were included.

2003). First year students enter an environment of 'mass-teaching', the enrolments in some first year subjects being in excess of 1000 students over the period of this study, having experienced considerably lower student-staff ratios in the final years of high school with a much higher flow of direct and ongoing feedback from instructors. In the university environment in the Department of Economics students have low contact hours with their teachers and are required to devote significant time to independent study.

We use three data sets. The first set of data is the average evaluation scores as reported for all subjects taught in the Economics Department from 2nd semester 1995 to 2nd semester 2003. This totals 565 observations. In appendix 1 a table is provided containing the core university questions (the first 9 questions) and the additional questions asked by the department in the QOT survey over the period 1995 – 2003. This table highlights which questions have remained constant and those that have changed over time. In particular those coded in the table 1, 2, 3, 5, 8, 9, 10, 11 and 12 have remained essentially the same over the period although they have varied in the order in which they have been asked in the QOT survey. In more recent years questions regarding multi media are now core questions.

Using this data we can assess the interrelationship between the question responses by using principal components analysis across the subjects taught over time. The first two components explain 63% of the variation in the questions that have been in the survey for the entire time frame of the data.[4] Table 1 provides correlations of the question responses (the factor loadings) with the first two components and these are plotted in Figure 1. From Table 1 it is clear that questions 1, 2, 3, 5, and 8 are highly correlated with the first component suggesting that all of these questions are eliciting the same information. Similarly, questions 11 and 12 are highly correlated with the second component whereas questions 9 and 10 are not very highly

---

[4] Since the 3rd component explains less than 10% we limit this analysis to the first 2.

correlated with either component. This relationship can also be seen from Figure 1 by observing the clustering of the questions.

Table 1: Correlation of responses to the first two components.

Question	1	2
2. Well taught	0.89	-0.25
8. Explanations by the instructor were clear	0.84	-0.09
3. Intellectually stimulating	0.80	-0.21
5. Teaching staff showed an interest	0.80	-0.21
1. I had a clear idea of what was expected	0.80	-0.17
11. Instructor is enthusiastic	0.58	0.69
12. My understanding (of economics) has improved	0.49	0.79
9. Workbook and/or handouts	0.44	-0.19
10. Recommended textbook(s)	0.20	0.34

Figure 1: Plot of the correlation of responses to the first two components
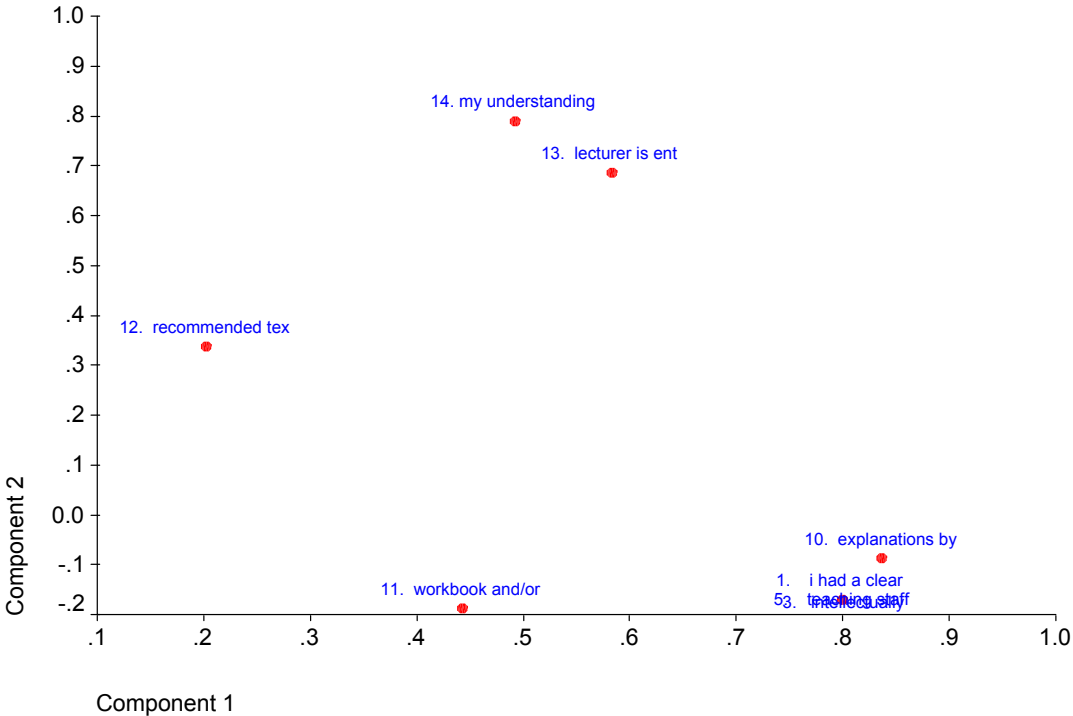
Brennan and Williams (2004) state that data from student evaluations will be more useful if they can be related to student profile data such as for example, age, gender, ethnic background and so on. This is because it could be used to determine how far views and experiences vary between types of students and if response rates are low to determine if responses are

representative of the student body. Although this information is not collected in the QOT survey at the University of Melbourne, we will derive an average student profile for each subject by using the student records for all those enrolled in the subject.

The second set of data is a student-specific data set consisting of student records for all students who have taken an economics subject from 1992 to 2003 (115,342 observations) which gave information on which subjects they took, their marks earned, and their rank in the subject. In addition, this data also included the characteristics of the student – their gender, their fee paying status, the secondary school they attended, the Faculty in which they are enrolled, their age, and their country of birth.

The third set of data is the characteristics of each subject - the year level of the subject (1st year, 2nd year etc), schedule times, class sizes, whether the subject was taught with more than one session, whether it was a quantitative subject (statistics and econometrics)[5], the semester the subject is taught in, the average response to the quality of teaching surveys administered in the subject, the response rate for the survey (proportion of number of surveys to the number enrolled) and the times when the subject evaluations took place.

Using these three data sources we constructed a subject/semester specific data set. The student subject evaluation survey data used starts in semester 2 1995 and finishes in semester 2 2003. The complete list of variables with descriptive statistics is reported in appendix 2. This amounts to 565 subjects taught by 95 different instructors with a total of 79,944 student enrolments and 53,835 QOT forms filled out.

To highlight some interesting features of the data set a number of profile plots are illustrated in Figure 2. Figure 2a shows the mean across the years of the "well taught" question (Q2) from the QOT survey over each semester. This figure highlights a semester difference in values. From 1997 until 2003 the mean values for semester 1 are much lower than for semester

---

[5] This Department provides a number of first and second year statistics and econometrics service subjects to the other departments in the Faculty of Economics and Commerce.

2. In addition, this plot also shows a large dip in the mean values for in semester 2 starting in 1999 and following through to 2001. Figures 2b and 2c illustrate the steady percentage increase in both Foreign and Australian full fee paying students. Figure 2d plots the average mark over all student enrolments for each semester over the time period. This plot highlights that until 2003 the average marks across student enrolments in 2nd semester are higher than in first semester again emphasizing there are semester differences. Figure 2e plots the proportion of QOT forms returned over all enrolments for each semester over the time period. In this figure we see that there has been a steady decline in the proportion of forms returned over time.

3.   Model Specification and Estimates

In this section we use a statistical methodology to identify how much of the variation in teaching evaluation scores are due to what happened in a subject and how much was predetermined by who was in the subject, how well they did, how large a subject it was and other factors that are not under the control of the instructor. For the analysis reported here we use ordinary least-squares estimation of a standard linear regression model. The regression specified is of the form:

$$\underbrace{Y_{ijt}}_{\substack{\text{Average} \\ \text{QOT score}}} = \underbrace{\alpha_i}_{\substack{\text{Fixed Effect} \\ \text{for lect } i}} + \underbrace{\gamma_t}_{\substack{\text{Year and} \\ \text{Semester}}} + \sum_{k=i}^{K} \underbrace{\beta_k X_{kjt}}_{\substack{\text{Subject} \\ \text{characteristics}}} + \varepsilon_{ijt} \qquad (1)$$

The average response for the well taught question (*Q2*) for each subject in the QOT survey is taken as the dependent variable. Table 1 and Figure 1 show that this question elicits basically the same information as a number of other questions in the survey and so our concentration has in fact a broader generality. Furthermore, although administrators at the University from which the data is drawn do give consideration to other questions, they tend to emphasise question 2.

The independent variables are the characteristics of the instructors, subjects and the students in the subject. The characteristics of the instructors are accounted for by including a set of dummy variables for each instructor in the data set.

The subject characteristics used include a dummy variable for each year level of the subject and in which semester it is in (*semester 1 1996, semester 1 1997...., semester 2 2003*), whether the subject is a quantitative subject or not (*quant*), class size (*enrol*), and the average responses to the questions on the usefulness of the recommended textbook (*Q9*) and the usefulness of the handouts/workbook (*Q10*)[6]. In addition, the class size variable is also interacted with the quantitative variable (*Quant*enrol*) and with the year level (*yr_1*enrol, yr_2*enrol, ..., yr_G*enrol*).

Of the variables measuring subject characteristics, none are in control of the instructor except, to some extent, the recommended textbook and the usefulness of the handouts/workbook. For these two variables, the influence of the lecturer is limited because the existence of a good textbook may depend on topics covered, and the latter is determined by the syllabus. Workbooks for some subjects are not entirely the product of the instructor because in some cases workbooks are handed down from previous instructors. The characteristics of the students in the classes are of two types. Among the first type are the demographic variables that are available from the student enrolment forms. These include the proportion of females (*female*), the proportions of students enrolled in different faculties across the University (*pf_econ, pf_art, pf_Eng, pf_sci, pf_law, pf_edu, pf_arc, pf_agr*), the proportion of students who are enrolled in a PhD program (not necessarily in the Economics program) (*pf_phd*), the proportions of students born in

---

[6] Other variables tried included class schedules and the day of the week the student survey was taken. Neither set of variables were significant and have not been included in the results reported here. Note also from Table 1 that Q9 and Q10 are quite distinct questions from the other questions in the survey.
[8] The University of Minnesota Website pertaining to the interpretation of student ratings of instruction (at http://www1.umn.edu/ohr/teachlearn/tasuper/interp.html) specifically states "Data from classes in which fewer than 75% of the students respond or in which there are fewer than 15 students may not provide reliable information".

Figure 2a: well taught
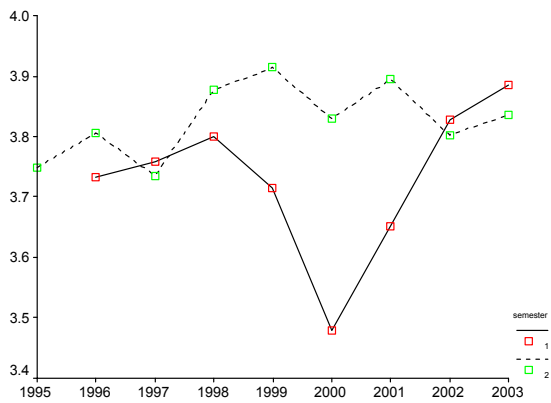
Figure 2d: Average mark over all enrolments.
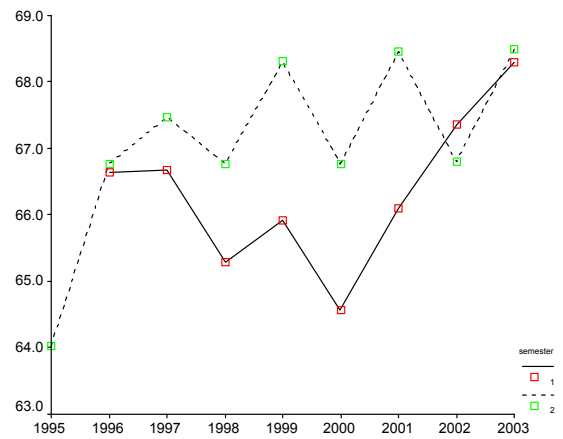
Figure 2b: % Foreign Fee over all enrolments
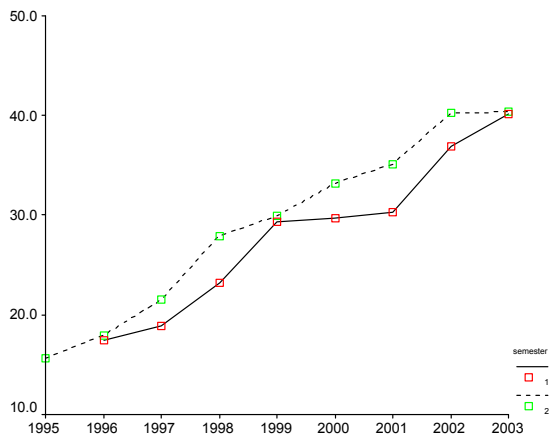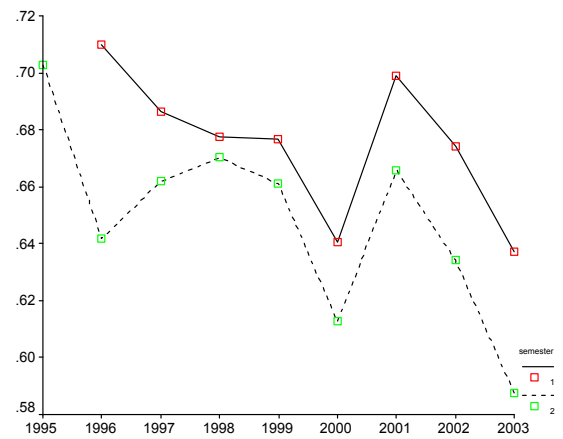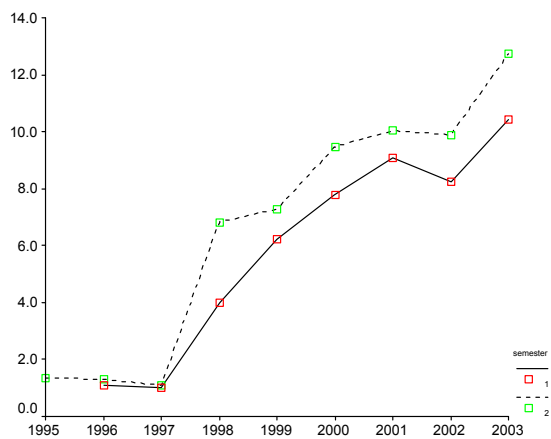

Figure 2e: Proportion of forms returned over all enrolments

Figure 2c: % Australian Fee over all enrolments

11

different countries (*cb_aus, cb_hk, cb_chn, cb_sng, cb_ind, cb_vn, cb_tia, cb_uk, cb_oth, cb_na, cb_isc, cb_weu*), the proportion of international and Australian students who are full fee paying and the proportion of Australian student with subsidized fees (*ft_AF, ft_OF, ft_AS*), if they are enrolled as honors students in the department or not (*hons*) and their average age (*avg_age*). We also have included the characteristics of the secondary school attended by the students in the subject. This is measured by the percent of the students that attended a government school (*ss_gov*), a catholic school (*ss_cath*) or an independent school (other denomination or private) (*ss_ind*) in Melbourne. If they attended a secondary school outside of Melbourne they were classified as *ss_nmelb*.

The second class of student characteristics indicates their status as students taking subjects in economics. These include their: average mark this year not including the current subject (*sem_mark*), the average difference between the mark received in the subject they are evaluating (something they do not know when making the evaluation) and their *sem_mark* (*d_smark*) and the average career response to the well taught question for the lecturers in the subjects taken this year excluding the current subject (*sem_q2*). Thus we first compute the average value for the "well taught" question in all the subjects these lecturers have taught for the entire period they have been in the Department, and then we use this average score as a guide to the quality of teaching in the subject they are teaching that year. Unlike the first group of student characteristics, the data used for these averages do not come from all the students in the subject. If the students have not taken or are not taking another Economics subject, we do not use their observation in the subject average that we are computing. Note this variable is not the average of all the QOT scores for the lecturers in the same semester or year excluding the subject under consideration because it is weighted by the particular schedules of the students in the class.

The purpose of including *sem_mark* is to condition the responses by the quality of the students enrolled in the subject. The *d_smark* variable is included so that we can measure a

"disappointment" effect. If we assume that the students have a good idea what their subject mark will be then *d_smark* measures the difference in the mark gained in this subject from their other subjects they are taking this year. One would anticipate if this subject gives them a higher grade relative to their other subjects they will be more pre disposed to considering this to be a well taught subject. The variable *sem_q2* is included to allow for the possibility that students complete student surveys by considering this subject relative to their rankings of instructors in the other subjects they are currently enrolled in and if it is a semester 2 subject the subjects they were also enrolled in during the first semester. Note we use the average QOT score for the instructor to represent his/her reputation rather than the particular score they received in a particular subject as a better measure of the overall teaching characteristics.

The QOT survey responses do not match the students in the class exactly because the survey is not filled out by all students who are enrolled or were enrolled in the subject. Students may have dropped the subject or failed to complete the survey. The response rate for the survey is determined by the number of students that attended the class in which the survey was administered. In particular, this means that for any subject there is a likelihood that the mean of the survey will not be representative of the students in the subject if the number of forms returned is too low. Figure 3 illustrates the response rates by the proportion of student enrolments in the Department of Economics subjects over the period 1996-2003. From Figure 3 it can be seen that over 11% of the students taught over this period were in subjects in which the response rate was less than 50% and more than 23% of the students were in subjects with a 53% or lower response rate. If we consider the 75% response rates for such surveys as proposed by some researchers we find that less than 23% of all the students taught were in subjects with a sufficient response rate to consider useful[8]. To account for response rates in the regression we followed the approach of Boex (2000) and included it as an explanatory variable.

Figure 3: Response rates by proportion of all students enrolled
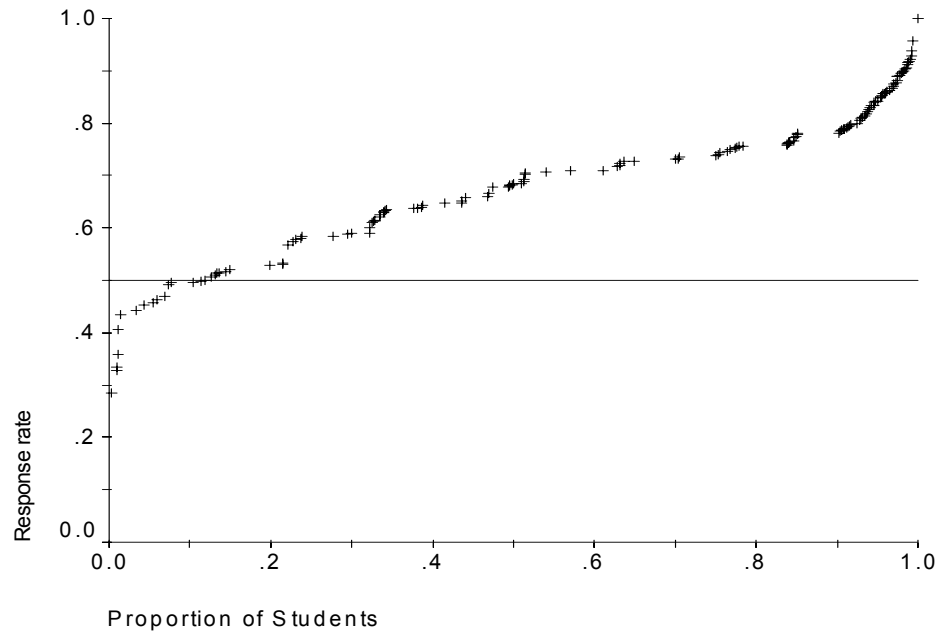
The regression results are reported in Table 2. The estimates of the standard errors in this model account for unobserved heteroskedasticity by grouping observations for each subject.[9] Individual instructor characteristics are taken account by including a dummy variable for each instructor in the data set. Table 2 reports the results of the estimate of 4 models. The first model includes the most complete set of regressors, in models 2, 3 and 4 the variables that record the proportion of students that attended secondary schools in the Melbourne area are dropped (*ss_ind, ss_gov* and *ss_cath*). Model 3 also drops the student specific variables based on marks and the comparative QOT score from the lecturers in other subjects – under the logic that this uses variables not known by the students when they are filling out the survey. Model 4 includes (*sem_mark, d_smark, sem_q2*) but excludes the two variables (the responses for textbook (*Q10*) and for workbooks (*Q9*)) that were also elicited by the survey and thus might be considered to be simultaneously determined and thus inducing potential bias in the estimated parameters.

---

[9] Thus all the observations for the same subject were used to estimate a separate variance which was then used to construct a "sandwich" or corrected covariance matrix for the least squares regression coefficients (see White 1980).

Variable	Model 1		Model 2		Model 3		Model 4	
	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat
Intercept	3.8775	4.56	3.8460	4.96	3.1706	6.13	4.2466	5.49
female	-0.0036	-2.10	-0.0035	-2.02	-0.0037	-2.17	-0.0027	-1.50
avg_age	0.0075	0.36	0.0079	0.38	0.0058	0.31	0.0297	1.39
ft_AF	0.0010	0.43	0.0010	0.44	0.0008	0.35	-0.0001	-0.07
ft_OF	0.0015	0.47	0.0020	0.72	-0.0002	-0.08	0.0027	0.91
RR	0.1333	0.86	0.1258	0.80	0.0617	0.38	0.2679	1.66
hons	0.3909	2.14	0.3685	2.01	0.4014	2.13	0.2735	1.28
ss_cath	0.0005	0.13						
ss_gov	-0.0013	-0.51						
ss_ind	-0.0004	-0.16						
d_smark	0.0107	2.13	0.0106	2.15			0.0110	2.07
sem_mark	0.0064	0.94	0.0064	0.94			0.0111	1.51
sem_q2	-0.3357	-3.59	-0.3355	-3.59			-0.4057	-4.20
Q10	0.0665	3.74	0.0664	3.71	0.0728	4.02		
Q9	0.1773	4.57	0.1789	4.69	0.1812	4.48		
ms	0.1726	1.18	0.1709	1.16	0.0659	0.37	0.1701	1.07
quant	0.1820	2.27	0.1820	2.25	0.1673	2.07	0.1794	2.05
enrol*quant	-0.0007	-3.20	-0.0007	-3.16	-0.0007	-3.19	-0.0009	-3.58
yr_1	-0.2487	-1.27	-0.2598	-1.31	-0.1839	-0.90	-0.1609	-0.74
yr_2	-0.0715	-0.42	-0.0910	-0.56	-0.0114	-0.07	-0.1199	-0.65
yr_3	0.1076	0.77	0.0887	0.66	0.1732	1.28	0.0576	0.38
yr_4	0.0421	0.33	0.0397	0.31	0.0227	0.17	0.1142	0.80
enrol*yr_1	-0.0001	-0.57	-0.0001	-0.61	-0.0001	-0.56	-0.0001	-0.59
enrol*yr_2	-0.0005	-2.09	-0.0005	-2.06	-0.0003	-1.10	-0.0004	-1.71
enrol*yr_3	-0.0013	-2.08	-0.0013	-2.09	-0.0016	-2.59	-0.0009	-1.40
enrol*yr_4	-0.0128	-3.22	-0.0126	-3.21	-0.0133	-3.15	-0.0143	-3.57
enrol*yr_G	-0.0064	-2.46	-0.0063	-2.45	-0.0071	-3.01	-0.0064	-2.13
cb_chn	0.0012	0.39	0.0007	0.24	0.0024	0.77	0.0012	0.36
cb_hk	-0.0158	-3.06	-0.0157	-3.10	-0.0184	-3.35	-0.0128	-2.46
cb_ind	0.0047	0.92	0.0042	0.83	0.0077	1.64	0.0038	0.64
cb_isc	0.0060	0.91	0.0059	0.89	0.0056	0.82	0.0040	0.61
cb_mal	0.0055	1.86	0.0052	1.77	0.0072	2.59	0.0027	0.94
cb_na	-0.0006	-0.11	-0.0008	-0.15	-0.0016	-0.29	0.0008	0.14
cb_oth	0.0010	0.29	0.0007	0.21	0.0018	0.51	0.0021	0.59
cb_sng	-0.0095	-2.62	-0.0098	-2.71	-0.0082	-2.28	-0.0112	-2.94
cb_tia	-0.0150	-0.65	-0.0159	-0.70	-0.0128	-0.63	-0.0221	-0.96
cb_uk	0.0275	3.49	0.0271	3.54	0.0259	3.26	0.0344	4.06
cb_vn	-0.0111	-1.01	-0.0111	-1.04	-0.0076	-0.78	-0.0160	-1.38
cb_weu	0.0072	3.74	0.0073	3.71	0.0077	3.91	0.0048	2.23
pf_Eng	-0.0053	-1.21	-0.0054	-1.22	-0.0053	-1.24	-0.0055	-1.18
pf_agr	0.0097	1.12	0.0092	1.07	0.0080	0.91	0.0162	1.70
pf_arc	0.0021	0.17	0.0013	0.11	0.0037	0.27	0.0037	0.28
pf_art	0.0024	1.42	0.0023	1.35	0.0027	1.73	0.0024	1.35
pf_edu	0.0013	0.75	0.0011	0.69	-0.0008	-0.42	0.0007	0.36
pf_law	0.0034	0.85	0.0034	0.85	0.0006	0.15	0.0045	1.01
pf_phd	-0.0011	-0.54	-0.0010	-0.49	-0.0005	-0.28	-0.0026	-1.30
pf_sci	-0.0026	-0.84	-0.0029	-0.99	-0.0043	-1.35	0.0001	0.02
semester 1 1996	-0.1089	-0.83	-0.1068	-0.84	-0.1590	-1.20	-0.2174	-1.69
semester 1 1997	-0.1699	-1.72	-0.1697	-1.73	-0.1612	-1.61	-0.3165	-3.15
semester 1 1998	0.0076	0.07	0.0055	0.05	-0.0306	-0.26	-0.2149	-2.00

---

[10] Note that the estimated coefficients and statistics on 95 dummy variables corresponding to individual instructor dummies have been included in the regression but have not been reported in the table. In the case when multiple dummy variables were included the excluded dummy variables include the 95th Instructor, semester 2 2003 and yr_G. Other groups of variables included were either proportions that sum to 1 or percentages that sum to 100. The excluded variables corresponding to these groups include pf_eco, cb_aus, ss_nmelb, and ft_AS.

Variable	Model 1		Model 2		Model 3		Model 4	
	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat
semester 1 1999	-0.1691	-1.71	-0.1663	-1.69	-0.1894	-1.83	-0.3621	-3.46
semester 1 2000	-0.2512	-2.79	-0.2548	-2.83	-0.2310	-2.48	-0.5086	-5.61
semester 1 2001	-0.1938	-2.09	-0.1916	-2.07	-0.2066	-2.04	-0.3609	-4.10
semester 1 2002	-0.1001	-1.10	-0.0998	-1.11	-0.0316	-0.33	-0.3631	-4.56
semester 1 2003	0.1793	1.71	0.1796	1.71	0.1877	1.81	-0.1814	-1.98
semester 2 1995	-0.2347	-2.02	-0.2272	-1.96	-0.2786	-2.36	-0.3537	-2.70
semester 2 1996	-0.0355	-0.33	-0.0330	-0.31	-0.0728	-0.66	-0.1847	-1.82
semester 2 1997	-0.0818	-0.80	-0.0820	-0.82	-0.1027	-1.01	-0.2432	-2.28
semester 2 1998	0.0057	0.07	0.0092	0.11	-0.0182	-0.20	-0.1192	-1.45
semester 2 1999	0.0186	0.19	0.0194	0.20	-0.0299	-0.31	-0.1820	-1.98
semester 2 2000	-0.1648	-1.77	-0.1640	-1.78	-0.1273	-1.34	-0.2796	-3.00
semester 2 2001	-0.0125	-0.14	-0.0092	-0.10	0.0151	0.18	-0.1850	-2.23
semester 2 2002	-0.0132	-0.17	-0.0084	-0.11	0.0427	0.55	-0.1634	-2.68

Most of the parameter estimates in Table 2 appear to be fairly robust to the choice of model specifications. From the results in Table 2 we see the subject characteristics that are important include semester and year level, class size and if a subject is quantitative. In particular, larger class sizes in 2nd, 3rd, 4th and graduate classes are negatively significant, while in 1st year subject enrolment is not significant. The lack of variation in class size for first year subjects is probably the cause for the low level of significance for this variable. The quantitative variable is positive although when interacted with class size it is significantly negative implying that it is larger sized quantitative classes that are associated with lower QOT scores. From these coefficients we can determine that if the class has a class size greater than approximately 260 students in a quantitative subject will have a lower QOT score than the non-quantitative subject of the same size. For Model 1 by the time these subjects get to 500 students (the maximum size of any lecture theatre used at Melbourne) the QOT score is decreased by .17 points. To account for multiple sessions we have included a dummy variable (*ms*) which is not found to be significant in any case.

Turning to student characteristics, we find that *d_smark* – the average of the difference between the average for all other subjects taken this semester and this subject across all students who took another subject this semester - is positive and significant over all specifications. These results indicate that the greater the relative mark, the higher the

students rate the instructor. However, we also find that the average of the other marks in other subjects does not have an influence on the QOT score.

We find that the coefficient estimated for *sem_q2*, the average q2 score for the other subjects that are taken this year, is negative and significant for all the specifications. This implies that the better the average performance of instructors in the other subjects taken by the students the lower the students rate the instructor in this subject. Note that this variable is a weighted average of the experiences of the all the students taking subjects for which we have QOT scores.

At the University of Melbourne there are large differences between faculties in the scores on question 2, with the Arts faculty generally receiving higher scores than professional faculties. This may reflect a difference in the expectations of students in different faculties. However, in general, the percentage of students from particular faculties do not show a significant effect, that is students from faculties other than Economics and Commerce do not rate economics instructors differently from Economics and Commerce students.

Counter to some other studies, except for Model 4, the proportion of females enrolled in the subject is found to have a negative and significant influence. Also counter to other studies, age of student has no significant effect. Note that because age will be highly related to other factors in this model - most foreign students and post-graduate students are older – the influence of age may not be measured by this model.

Somewhat surprisingly, the requirement to pay fees apparently has no significant impact. We find insufficient evidence that fee-paying students are more demanding of their instructors. Note that this follows for both foreign fee paying students as well as domestic fee paying students.

We also find that the percent of students that attend independent secondary schools (requiring private tuition that in some cases is more than twice the cost of university), or catholic schools or government schools, apparently has no impact on the QOT score. These

variables were included to determine if these students had higher expectations based on the greater level of resources in their secondary school. Again this result may have to do with the composition of the students in this Department. From Appendix 2 we find that on average 47% of the students in the subjects we consider here did not attend a local secondary school (*ss_nmelb*).

One characteristic of the students that does appear to be important is their status as honours students in the Economics Department. Honours students are able to take 3rd year subjects and anecdotally it had been proposed that they are more "demanding" than the other students in these subjects. However, we find evidence of the opposite that the higher the proportion of honours students in a subject (*hons*) the higher the QOT all other things taken into consideration. One possible interpretation of this result is that the honours students are attracted to subjects taught by instructors that they have heard are "good teachers".

The response to the question concerning the quality of the workbook and handouts (*Q9*) and the textbook (*Q10*) are both highly positively significant with the magnitude of the coefficient for the workbook almost a factor of three times the coefficient on the textbook. In line with our comment above about the limited control that instructors have over the quality of the workbook and textbook, these estimates suggest a role of fortune in being supplied with good material. However, instructors do have some control and in as far as they prepare a workbook, select a textbook, or even write their own textbook, these estimates suggest that they will be rewarded with higher teaching evaluations. To allow for this possibility that good teaching implies the use of good materials for the subject we estimate Model 4 to demonstrate that most of the results we have discussed for other variables are not influenced by the presence of these variables in the other regressions.

The coefficients for the country of origin reveal an interesting pattern. The percent of students whose country of birth is Hong Kong or Singapore is negatively significant whereas the proportion of students whose country of birth is the UK or Western Europe is positively

significant. This suggests a contrast between East Asia and European backgrounds, with the former being more critical of their instructors. However this inference appears to fail when we note that for Model 3 and to a lesser extent Models 1 and 2, the percent of Malaysian born students has a positive influence, the insignificance of the other East Asia dummies, that is: China, Indonesia, Vietnam and Taiwan, suggests the influence of more than simply an east Asia-Europe contrast. This may indicate that expectations are different for those students that come from a more affluent Asian country.

The response rate (RR) is not significant; however this is a difficult parameter to interpret. The response rate may be considered a measure of lecture attendance on the day the survey is administered. The QOT Survey in the Economics Department at Melbourne is usually administered during the penultimate week of the semester which is not atypical session for the semester. If we assume that it is a proxy for attendance then low response rates may indicate that students have either given up on the lectures or that the other sources for information about the lecture are so complete (via recordings and lecture slides posted on the subject web-site) that the students feel it is unnecessary to attend the lecture.

The F - statistics computed to test groups of coefficients based on Model 1 are reported in Table 3. These results indicate that we can reject the hypothesis that lecturers are the same, that the level of the subject, that the enrolment in the subject, that the year and semesters are all the same, and that country of birth does not matter. In addition, we can reject the hypothesis that all the coefficients for the quantitative dummy and the interaction with enrolment are zero, as well as that the coefficients for the influence of the workbooks and the textbook are zero, that the student's in the subject's semester effects (their marks, and QOT experience in other subjects) is equal to zero. However we are unable to reject that the proportion of students from different Faculties, the proportion of students by fee paying status, and the proportion of students by type of secondary school have any influence on QOT scores.

Table 3: Results on *F*-statistics

| Aggregate Effects	*F*-stat	DF$_1$[11]	*p*-value
Lecturers	2739.89	95	0.0000
Gender & age	2.31	2	0.1016
Fee Status	0.19	2	0.8302
School Type	0.14	3	0.9333
sem_mark, d_smark, sem_q2	5.56	3	0.0010
Books and Workbooks	16.01	2	0.0000
Quantitative	5.20	2	0.0061
Level	4.72	10	0.0000
Enrolment	5.46	6	0.0000
Country of Birth	4.72	12	0.0000
Faculty	0.81	8	0.5905
*Semester*year*	2.45	16	0.0017

4. Evaluating instructor performance

To assess the quantitative importance of variables outside the instructor's control on their teaching evaluations, we calculated from these regression results a conditional mean for each instructor where the subject and average student conditions are accounted for. To determine how these factors may influence the average score for question 2 across all subjects taught by an instructor, in Figure 4 we plot the raw average by instructor on the horizontal axis and the conditioned average for the same instructor on the vertical axis. In order to compute the conditioned average we predict the value of the QOT score for each instructor under the assumption that they teach a subject with average characteristics and an equal chance of teaching in any year and quarter in the sample. This is done by assuming the relationship below holds:

$$\underbrace{\hat{Y}_i}_{\substack{\text{Conditional} \\ \text{QOT score}}} = \underbrace{\hat{\alpha}_i}_{\substack{\text{Fixed Effect} \\ \text{for lect i}}} + \underbrace{\sum_{t=1}^{T} \frac{1}{T} \hat{\gamma}_t}_{\substack{\text{Year and} \\ \text{Semester}}} + \underbrace{\sum_{k=i}^{K} \hat{\beta}_k X_{kjt}}_{\substack{\text{Subject} \\ \text{characteristics}}} \qquad (2)$$

The assumption being made here is that the conditioned average removes the influence of the class and student specific factors. For the results discussed here we use the specification in Model 1; however any of the specifications could be used.

---

[11] All the denominator degrees of freedom for these tests is 284.

Figure 4: Raw Average vs. Conditional Average QOT Scores
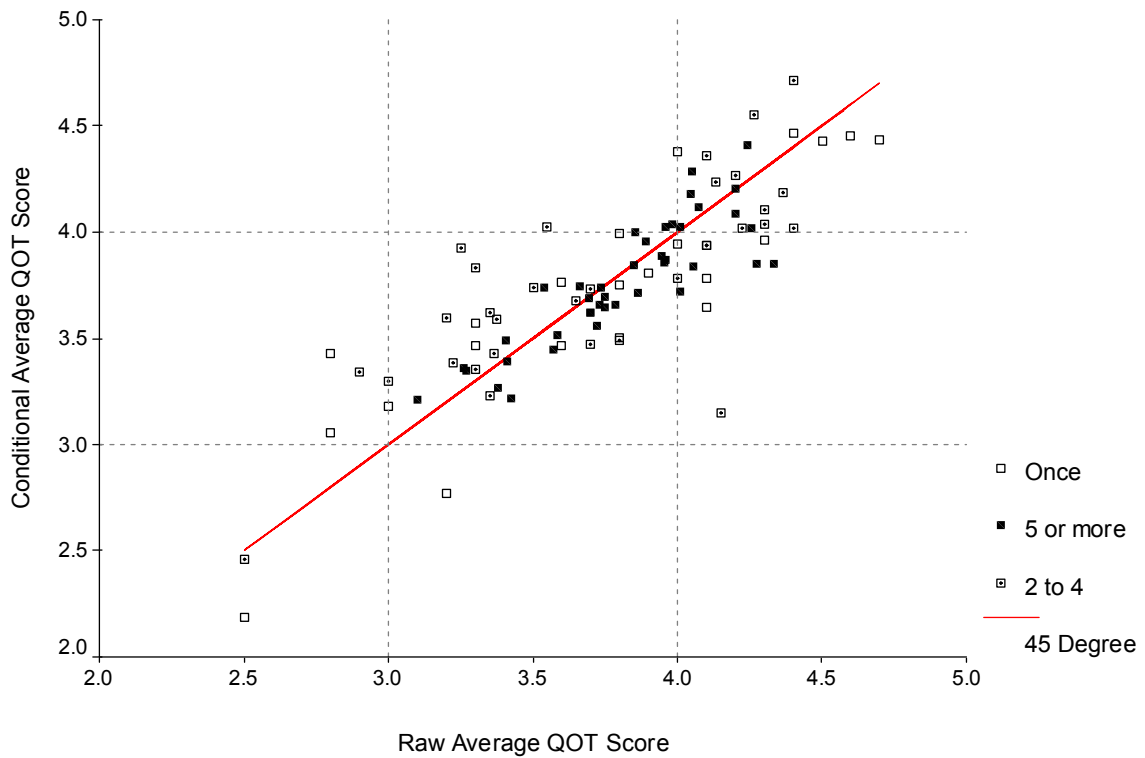
In Figure 4 the points that lie on the 45° line indicate that the class and student specific factors have little influence on the difference between the conditioned and unconditioned QOT scores. Observations above the 45° degree line indicate that there are instructors for whom the conditional averages are better than their unconditional averages, or raw scores. This implies that the raw scores under-value the teaching performance of the instructor. The converse is true for observations below the 45° line.

At the University of Melbourne, a score of less than three is thought to be a clear signal of poor teaching, though the difference between, say 3.5 and 3.8 is thought to be less informative. In Figure 4 it can be seen that five lecturers received an average raw QOT score for all the subjects they taught of three or less and yet their adjusted scores were greater than three. In one case the adjustment raise the score from 2.8 to 3.5. Thus we find that the use of such arbitrary yardstick values may be misleading.

In the data set there are instructors with a wide range of teaching experience within the department. There are those that have only taught one subject and others who have

taught more that 20 subjects over this period. To see if there are distinguishing characteristics for those that fell below the 45° line as opposed to those that fell above, the symbol used to plot the observation indicates the number of subjects they have taught in the Economics Department since the QOT scores began. From Figure 4 we can also see that instructors who have only taught one subject are more likely to be either way above the 45° line or way below the line. Instructors who have taught two-four subjects are still more likely to be above or below the 45° line but not by as much. Instructors who have taught more than five subjects are much more likely to be an observation along the 45° line. It would therefore seem sensible to look at an average of overall QOT scores after instructors have taught a mix of subjects rather than simply concentrating on one particular value. This approach seems to give similar results as to the approach that conditions QOT scores on subject and average student characteristics.

5.   Conclusions

In this analysis we have demonstrated that the characteristics of the subject and students enrolled in the subject have a significant impact on the quality of teaching scores obtained for a subject. In most cases, these characteristics are not under the control of the instructor and consequently the quality of teaching scores should be adjusted to account for these conditions. This is particularly important where these scores are being used to inform appraisal outcomes and tenure and promotion decisions across individuals who teach subjects with different characteristics.

We found that class size impacts negatively on student QOT ratings in 3rd and 4th year subjects but, with the exception of quantitative subjects, is not significant at first and second year. Previous research indicates that mathematics/technical subjects tend to score lower than more discursive subjects (see for example Braskamp & Ory, 1994; Neumann, 2000; Sixbury & Cashin, 1995a, 1995b). We found that quantitative subjects with class sizes less than 260 rated marginally more positively than other subjects. Nevertheless, for

quantitative subjects where class sizes exceeded 260 there was a significant negative relationship as anticipated. Consequently class size can be seen as a source of bias in the student evaluation of quality of teaching where the subject is quantitative and where the subject is taught in the latter years of the undergraduate degree.

Some studies (Mason, Steagall and Fabritus 1995) indicate that females are more likely to give positive ratings of teacher effectiveness than males however our study indicates that the proportion of females in a class leads to marginally negatively significant QOT scores.

Recommended textbooks and handouts and workbooks have a significant positive influence on QOT scores. Both of these variables are not necessarily in the control of the instructor. The selection of a good text depends on the topics covered and often instructors use or modify subject materials developed from the previous instructor of the subject. However, this result may indicate that instructors would be advised to choose textbooks and develop subject materials wisely and not proceed with the status quo.

Of particular interest in light of the increasing proportion of international students in Australian universities, is the impact of country of birth on student evaluations of teaching. Our study indicates that in subjects where a high proportion of students are from Hong Kong or Singapore QOT scores are likely to be lower. Conversely if there is a higher proportion of students from the UK, Western Europe and in some cases Malaysia, QOT scores will be higher. However, for a number of countries from which a significant proportion of students originate (China, Indonesia, and Indian sub-continent) there is no significant impact on QOT scores. This indicates that student learning cultures and expectations differ among students from Asian countries. In particular, this result may highlight that students from Hong Kong and Singapore have specific demands which will have implications for teaching practice.

We find that the higher the average mark of students in the subject as well as the improvement in their mark the higher they will score the instructor. However, there does not

appear to be a marked secular trend in grades (See Figure 2d). We also identified the "relative effect" whereby QOT scores given to instructors in other subjects taken by students during the same year is negatively significant and provides an indication of some cross subject effects that are not in the control of the instructor. Students seem to complete surveys in light of their experience with other instructors in the subjects they are enrolled in that year.

We find that by estimating this relationship we are able to condition the average QOT scores in order to account for these exogenous factors. When this is done we find that in some cases the difference between the raw average QOT score and the conditioned score is over 1 point and often enough to move scores over and under any particular score that may be used for evaluative purposes. For example, there are several cases in our data set where setting the yardstick for poor teaching at a score less than 3 appears to give an unreasonable judgement that the teaching was poor.

We also note that the adjustment of raw scores is greatest for instructors who have taught the fewest times. This implies that the value of the average of QOT scores of instructors that have taught a wider variety of subject types is close to the value obtained conditioned on the characteristics of the subject and is therefore a more reliable signal of teaching performance. However, this does not imply that a single subject QOT evaluation for a more experienced instructor is of any more reliance than for a first time instructor.

Of general concern is the steady decline in the number of responses to the survey over the period from 1995-2003 (Figure 2e). Student weariness with completing survey forms for every subject, every semester may be alleviated to some extent through more open communication of the results and strategies put in place to address their concerns.

While not wishing to underestimate the effect of instructor characteristics (such as enthusiasm, clarity, showing good management of student behaviour, demonstrating well developed interpersonal skills, being able to provide intellectual stimulation, showing respect

for students, being organised and having good presentations skills) on student perceptions of teaching our study has shown that there are many exogenous factors that also need to be considered when making judgments about an individuals teaching performance. Heads of departments, and tenure and promotion committees should be carefully informed in relation to the interpretation of quality of teaching scores and should avoid the simplistic approach of comparing means when making decisions.

Appendix 1:   QOT questions 1995-2003

code	1995	1996	1997	1998 to 1999	2000 to 2003
Q1	1. The aims of the subject were made clear.	1. The aims of the subject were made clear.	1. I had a clear idea of what was expected of me in this subject.	1. I had a clear idea of what was expected of me in this subject.	1. I had a clear idea of what was expected of me in this subject.
Q2	2. This subject was well taught.	2. This subject was well taught.	2. This subject was well taught.	2. This subject was well taught.	2. This subject was well taught.
Q3	3. This subject was intellectually stimulating.	3. This subject was intellectually stimulating.	3. This subject was intellectually stimulating.	3. This subject was intellectually stimulating.	3. This subject was intellectually stimulating.
Q4	4. This subject helped develop my learning skills.	4. This subject helped develop my learning skills.			
Q4a			4. I received helpful feedback on how I was going in this subject.	4. I received helpful feedback on how I was going in this subject.	4. I received helpful feedback on how I was going in this subject.
Q5	5. In this subject, teaching staff showed an interest in the academic needs of students.	5. In this subject, teaching staff showed an interest in the academic needs of students.	5. In this subject, teaching staff showed an interest in the academic needs of students.	5. In this subject, teaching staff showed an interest in the academic needs of students.	5. In this subject, teaching staff showed an interest in the academic needs of students.
Q6	6. The outcomes of previous questionnaires in this subject have been explained to me in a constructive manner.	6. The outcomes of previous questionnaires in this subject have been explained to me in a constructive manner.			
Q6a			6. The volume of work in this subject was appropriate.	6. The volume of work in this subject was appropriate.	6. The volume of work in this subject was appropriate.
Q7	7. The lectures were well prepared.				
Q7i				7. The multimedia based technology helped me to learn effectively.	
Q7a					7a. The teaching in this subject regularly included computer-based multimedia programs. (yes/No)
Q7b					7b. The computer-based multimedia

code	1995	1996	1997	1998 to 1999	2000 to 2003
					programs helped me to learn effectively.
Q8	8. The explanations by the instructor of concepts and details were clear.	7. The explanations by the instructor of concepts and details were clear.	7. The explanations by the instructor of concepts and details were clear.	9. The explanations by the instructor of concepts and details were clear.	10. The explanations by the instructor were clear.
Q8i				8. I regularly made use of the information and materials made available by the teaching staff on the Internet.	
Q8a					8a. My learning activities in the subject regularly made use of the web. (Yes/No)
Q8b					8b. Using the web helped me to learn effectively.
Q9	9. The workbook/handouts have been a useful guide for the subject.	8. The workbook and/or handouts have enhanced my learning in this subject.	8. The workbook and/or handouts have enhanced my learning in this subject.	10. The workbook and/or handouts have enhanced my learning in this subject.	11. The workbook and/or handouts have enhanced my learning in this subject.
Q9a					9. I found it useful to access information and subject materials through the subject website.
Q10	10. The recommended textbook(s) in this subject were useful.	9. The recommended textbook(s) in this subject were useful.	9. The recommended textbook(s) in this subject were useful.	11. The recommended textbook(s) in this subject were useful.	12. The recommended textbook(s) in this subject were useful.
Q11	11. The instructor is enthusiastic about this subject.	10. The instructor is enthusiastic about this subject.	10. The instructor is enthusiastic about this subject.	12. The instructor is enthusiastic about this subject.	13. The instructor is enthusiastic about this subject.
Q12	12. My understanding of this subject has been improved significantly.	11. My understanding of this subject has been improved significantly.	11. This subject has improved significantly my understanding of the discipline of economics.	13. This subject has improved significantly my understanding of the discipline of economics (actuarial sciences).	14. This subject has improved significantly my understanding of the discipline of economics (actuarial studies).
Q12a		12. In this subject, the grade I expect to get is: H1, H2, H3, P or Fail.	12. In this subject, the grade I expect to get is: H1, H2, H3, P or Fail.	14. In this subject, the grade I expect to get is: H1, H2, H3, P or Fail.	

Appendix 2: Variable Names, Descriptions and Sample Statistics by Subject

Variable	Label	N	Mean	Std Dev
N_FORMS	Number of forms returned	565	95.284	194.823
RR	% of form returned	564	0.731	0.161
Q1	1. I had a clear idea of what was exp	565	3.642	0.468
Q2	2. well taught.	565	3.790	0.563
Q3	3. intellectually stimulating.	564	3.811	0.502
Q4	4. helped develop my learning skills.	95	3.598	0.499
Q5	5. teaching staff showed an interest	564	3.830	0.520
Q6	6. outcomes of previous questionnaire	95	2.752	0.634
Q7	7. lectures were well prepared.	159	3.698	0.570
Q8	10. explanations by the instructor were c	565	3.736	0.600
Q9	11. workbook and/or handouts have enhanced	562	3.630	0.778
Q10	12. recommended textbook(s) in this sub	565	3.224	1.236
Q11	13. instructor is enthusiastic about this	564	3.900	0.893
Q12	14. subject has improved significantly	335	3.513	1.078
Q12A	12. In this subject, the grade I expect	247	71.929	5.910
Q4A	4. I received helpful feedback on how	469	3.263	0.556
Q6A	6. The volume of work in this subject	470	3.730	0.439
Q7I	7. The multimedia based technology he	124	2.075	1.768
Q8I	8. I regularly made use of the inform	124	1.819	1.600
Q7B	7b. The computer-based multimedia program	282	2.623	1.728
Q8_C	Useful Web Materials	366	2.867	1.312
Q8B	8a. Using the web helped me to learn e	281	3.196	1.486
Q9A	9. I found it useful to access inform	282	3.451	0.657
Q15	14. The computer facilities provided by	107	2.631	1.397
LEVEL	year level	565	3.913	1.910
SEM	1=1st sem, 2=2nd sem	565	1.579	0.494
YEAR	YEAR	565	1999.380	2.508
Yr_enrl	Yr of Enrol	565	1999.380	2.508
Enrol	Enrolment	565	141.494	282.013
yr_1	1st yr subject	565	0.090	0.287
yr_2	2nd yr subject	565	0.110	0.313
yr_3	3rd yr subject	565	0.336	0.473
yr_4	4th yr subject	565	0.166	0.373
yr_G	Graduate subject	565	0.297	0.457
Quant	A Quantitative subject in Economics	565	0.232	0.422
ms	Subject with multiple sessions (enrol > 500)	565	0.094	0.292
sem_mark	avg mark this yr (not this one)	547	70.971	5.815
d_smark	Average difference between sem_mark and current mark	547	1.453	5.430
hons	Proportion of Honours students in the subject	565	0.170	0.273
sem_q2	avg q2 score in subjects taken this yr (not this one)	547	3.460	0.630
cb_aus	Australia born	565	65.416	23.741
cb_mal	Malaysia born	565	6.774	9.262
cb_hk	Hong Kong born	565	2.425	3.804

Variable	Label	N	Mean	Std Dev
cb_chn	China born	565	4.790	10.336
cb_sng	Singapore born	565	4.478	6.304
cb_ind	Indonesia born	565	3.112	5.352
cb_vn	Viet Nam born	565	0.705	2.810
cb_tia	Taiwan born	565	0.438	1.202
cb_uk	UK born	565	0.424	2.091
cb_oth	Other country born	565	6.598	8.321
cb_na	North American born	565	2.013	4.762
cb_isc	Indian sub-cont born	565	1.341	3.262
cb_weu	Western EU born	565	1.486	5.207
avg_age	Average age when taking all subjects	565	23.806	3.249
Male	% male students	565	58.880	14.133
Female	% female students	565	41.120	14.133
pf_eco	% Economics Faculty	565	66.976	23.456
pf_Eng	% Engineering	565	4.349	6.443
pf_art	% Arts	565	7.282	17.270
pf_arc	% Architecture	565	0.159	1.200
pf_sci	% Science	565	6.933	7.953
pf_law	% Law	565	6.053	7.018
pf_edu	% Education	565	1.797	13.196
pf_agr	% Agriculture	565	0.536	1.392
pf_phd	% PhD	565	5.817	15.837
ft_AF	% Australian Fee	565	9.486	15.928
ft_OF	% Foreign Fee	565	28.826	20.019
ft_AS	% Australian Subs	565	61.688	26.789
ss_gov	% Government school	565	15.067	10.207
ss_ind	% Independent school	565	27.413	15.798
ss_cath	% Catholic school	565	10.031	7.328
ss_nmelb	% not from a Melbourne Secondary school	565	47.303	25.048

References

Abrami, P.C., L. Leventhal and R. P. Perry (1982), "Educational Seduction", Review *of Educational Research*, 52(3), 446-464.

Aigner, D. and F. Thum (1986), "On Student Evaluation of Teaching Ability", *Journal of Economic Education*, 17, 243-265.

Annual Report (2003), Department of Economics, University of Melbourne.

Basow, S.A. and N. T. Silberg (1987), "Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?", *Journal of Educational Psychology*, 79(3), 308-314.

Boex, L. F. J. (2000), "Attributes of Effective Economics Instructors: An Analysis of Student Evaluations", *Journal of Economic Education*, 31(3), 211-227.

Bosshardt, W. and M. Watts (2001), "Comparing Student and Instructor Evaluations of Teaching", *Journal of Economic Education*, 32(1), 3-17.

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

Brennan, J. and R. Williams (2004), "*Collecting and Using Student Feedback. A Guide To Good Practice*", Learning and Teaching Support Network, Centre for Higher Education Research and Information.

Casey, R. J., P. Gentile and S. Bigger (1997), "Teaching appraisal in higher education: an Australian perspective", *Higher Education, 34*(4), 459-482.

Costin, F., W. H. Greenough and R.J. Menges (1971), "Student Ratings of College Teaching: Reliability, Validity and Usefulness", *Review of Educational Research*, 8, 511-535.

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" at *http://www.brevard.edu/fyc/fya/CuseoLink.htm*.

Davies, M., J. Hirschberg, J. Lye, C. Johnston and I. McDonald (2005), "Is it your fault? Influences on student evaluations of teaching in tertiary institutions", available at *http://tlu.ecom.unimelb.edu.au/academic_resources/publications.html*

DeCanio, S. J. (1986), "Student Evaluations of Teaching – A Multinomial Logit Approach", *Journal of Economic Education*, 17(3), 165-176.

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.

Gokcekus, O. (2000), "How do University Students Value Economics Courses? A Hedonic Approach", *Applied Economics Letters*, 7, 493-496.

Greenwald, A. G. and G. M. Gillmore (1997), "No Pain, No Gain? The Importance of Measuring Course Workload in Student Ratings of Instruction", *Journal of Educational Psychology*, 89(4), 743-751.

Jacobs, L. C. (2002), *Student Ratings of College Teaching: What Research Has to Say?* Retrieved 10/5/ 2005, from http://www.indiana.edu/~best/multiop/ratings.shtml.

Kember, D., D. Leung and K. Kwan, (2002), "Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching?", *Assessment & Evaluation in Higher Education*, 27, 5, 411-425.

Liaw, S. H., and K. L. Goh (2002), "Biases in student evaluation of teaching: the case of Faculty of Economics & Administration University of Malaya", *FEA Working Paper, No. 2002-3*.

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

Marsh, H. W and M. Dunkin (1992), "Students' Evaluations of University Teaching: A Multidimensional Perspective", in J.C. Smart(ed), *Higher Education: Handbook on Theory and Research*, 143-234, New York: Agathon Press.

Marsh, H. W. and J. E. Jr. Ware (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.

Mason, P. M., J. W. Steagall and M. M. Fabritius (1995), "Student Evaluations of Faculty: A New Procedure for Using Aggregate Measures of Performance", *Economics of Education Review*, 14(4), 403-416.

McKeachie, W. J. (1978), *Teaching Tips: A Guidebook for the Beginning College Teacher*, Lexington MA: Heath.

Murray, H. G., J. P. Rushton and S. V. Paunonen (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. and K. A. Lynch, (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.

Seldin, P. (1993), "The Use and Abuse of student Ratings of Professors", *The Chronicle of Higher Education,* 46(6), 687-700.

Shevlin, M., P., Banyard M., Davies and M.Griffiths (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.and J. A. Siguaw (2000), "Student evaluations of teaching: an exploratory study of the faculty response", *Journal of Marketing Education, 22*(3), 199-213.

Sixbury, G. R. and W. E. Cashin (1995a), Comparative data by academic field. In *IDEA Technical Report, vol. 10*: Center for Faculty Evaluation and Development, Manhattan, Kansas: Kansas State University.

Sixbury, G. R. and W. E. Cashin (1995b), Description of database for the IDEA Diagnostic Form. In *IDEA Paper, vol. 9*: Center for Faculty Evaluation and Development, Manhattan, Kansas: Kansas State University.

Ting, K. (2000), "A Multi-Level Perspective on Student Ratings of Instruction", *Research in Higher Education*, 41(5), 637-661.

White, Hal (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.

Williams, R. and J. C. Ory (1992), "*A Further Look at Class Size, Discipline Differences and Student Ratings*", Office of Instructional Resources, University of Illinois, Urbana-Champaign.

Worthington, A. C. (2002), "The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A Case Study in Finance Education", *Assessment and Evaluation in Higher Education*, 27(1), 49-64.