ISSN 1444-8890

ECONOMIC THEORY, APPLICATIONS AND ISSUES

Working Paper No. 42

Students' Evaluation of Teaching Effectiveness: What Surveys Tell and What They Do Not Tell

by

Mohammad Alauddin
And
Clem Tisdell

November 2006

THE UNIVERSITY OF QUEENSLAND

ISSN 1444-8890
ECONOMIC THEORY, APPLICATIONS AND ISSUES
(Working Paper)

Working Paper No. 42

Students' Evaluation of Teaching Effectiveness: What Surveys Tell and What They Do Not Tell

by

Mohammad Alauddin[1]
and Clem Tisdell [*]

November 2006

[1] School of Economics, The Univesity of Queensland, Brisbane 4072 Australia.
Tel: +61 7 3365 6570 Fax: +61 7 3365 7299 Email: m.alauddin@economics.uq.edu.au.

[*] School of Economics, The University of Queensland, Brisbane 4072 Australia.
Tel: +61 7 3365 6570 Fax: +61 7 3365 7299 Email: c.tisdell@economics.uq.edu.au

Students' Evaluations of Teaching Effectiveness: What Surveys Tell and What They Do Not Tell

ABSTRACT

Employing student evaluation of teaching (SET) data on a range of undergraduate and postgraduate economics courses, this paper uses ordered probit analysis to (i) investigate how student's perceptions of 'teaching quality' (TEVAL) are influenced by their perceptions of their instructor's attributes relating including presentation and explanation of lecture material, and organization of the instruction process; (ii) identify differences in the sensitivty of perceived teaching quality scores to variations in the independepent variables; (iii) investigate whether systematic differences in TEVAL scores occur for different levels of courses; and (iv) examine whether the SET data can provide a useful measure of teaching quality.

It reveals that student's perceptions of instructor's improvement in organization, presentation and explanation, impact positively on students' perceptions of teaching effectiveness. The converse appears to hold. The impacts of these factors vary between postgraduate and undergraduate pograms as well as between levels within the undergraduate program. The pragmatic implications of SET procedures are discussed. It is argued that while they are simple to apply, there are dangers of using them to judge the quality of teaching. From a practical point of view, they are a poor indicator of teaching performance and in themselves provide no guidance to lecturers as to how to improve their teaching performance.

Key words: Teaching effectiveness, Instructor attributes, Ordered probit, Sensitivity analysis, Underdetermination, Pseudoscience

JEL Classification: A2, I2.

Students' Evaluations of Teaching Effectiveness: What Surveys Tell and What They Do Not Tell

1. Introduction

Student evaluation of teaching is a burgeoning industry, which has witnessed spectacular growth over the last three decades or so. As Wilson (1998, p.A12) states '... Only about 30 per cent of colleges and universities asked students to evaluate professors in 1973, but it is hard to find an institution that doesn't today. And student ratings carry more and more weight. ... Such evaluations are now the most important, and sometimes the sole, measure of an instructor's teaching ability'. It is invariably used in promotion or tenure decisions as the most important indicator of teaching 'quality'. This notwithstanding, there is considerable controversy surrounding the derivation and use of teaching effectiveness instruments[1]. This paper argues that despite their widespread use, student evaluations of teaching involve an inexact science. As Mason, Steagall and Fabritius (1995, p.403) rightly put it:

> '...Students are not fully informed consumers because they do not necessarily know whether the professor is providing them with the relevant material, and doing so correctly. Consequently, students' judgment may be insufficiently well informed to evaluate this portion of the performance of their professors. Furthermore, students may not be fully cognizant of the quality until later life experiences dictate the long-term value transferred.[2] In addition, the methodological approaches employed by a professor may be effective for a particular student, or even the majority of students, but they are unlikely to the best for all of the students. Student evaluation scores will reflect both the views of those students for which (sic) the methods work, and those for which (sic) they do not'. …'.

In conformity with a large body of literature, this paper assumes that teaching is a multidimensional process in that '...an instructor's overall effectiveness depends on these instructor attributes, such as the clarity of the instructor's lectures, the course organization, the degree to which the instructor motivates students, and the instructor's success in building an interpersonal rapport with students' (Boex 2000, p.211).[3]

This paper uses a large sample of student evaluation data on teaching (SET data) and seeks answers to the following questions:

- What are the principal determinants of teaching effectiveness score?
- How does an increase or decrease in the perceived score of any determinant affect the probability of a higher or lower score for perceived teaching effectiveness?
- Do the impacts of these factors vary between postgraduate and undergraduate programs, and between levels within the undergraduate program?
- Do or can SET data provide useful measures of teaching effectiveness or quality?
- Do they provide guidance on how teaching quality can be improved?

Studies employing econometric investigation of the effects of instructional attributes of teaching and learning are few in the existing literature on economics education (DeCanio 1986; Mason et al. 1995; Boex 2000). Nevertheless, in our view, they represent an advance over the education and/or educational psychology literature. The existing education literature, dominated principally by the educational psychology literature, implicitly assumes, almost as an article of faith, that an average student *inter alia* allocates the expected number of hours to study, comes well prepared for tutorials/lab sessions, consults the teaching staff on a regular basis, and does not leave all or most of his/her studies until very late in the semester/term. That these variables/attributes determine student attitude and behavior toward learning can affect students perception's of teaching effectiveness has barely been addressed in the existing literature[4].

This paper is organized as follows: Section 2 presents the theoretical model. Section 3 outlines the main features of the data. Section 4 presents and discusses the empirical results. Section 5 provides and examines results of sensitivity analysis. Section 6 presents an analysis of pragmatic implications of SET procedure. Section 7 presents a concluding overview and comments.

2 Methodology: The Ordered Probit Model

A large body of literature recognizes that linear regression is inappropriate when the dependent variable is categorical, especially if it is qualitative[5]. The appropriate theoretical model in such a situation is the ordered probit model (see for example, Greene 2000, pp.875-79). Over the last three decades or so these models have been widely used as a methodological framework for analyzing ordered data since the pioneering work of McKelvey and Zovoina (1975). In contrast to most of the education or educational psychology literature, the economics education literature uses ordered probit and/or multinomial logit models (DeCanio 1986, Mason et al. 1995; Boex 2000; Chan, Miller and Teha 2005).

Consider the following model which is built around a latent regression

$$y^* = x'\beta + \varepsilon \tag{1}$$

Where y* is unobserved. What is observable is:

$$\begin{aligned} y &= 0 \ \text{ if } y^* \le 0 \\ &= 1 \text{ if } 0 < y^* \le \mu_1 \\ &= 2 \text{ if } \mu_1 < y^* \le \mu_2 \\ &\ldots\ldots \\ &= J \ \text{ if } y^* \ge \mu_{J-1} \end{aligned} \tag{2}$$

The μ's are unknown threshold parameters to be estimated with β. Thresholds parameters determine the estimations for different observed value of y. These threshold parameters can be interpreted as intercepts in equation (1).

Consider, for example, an opinion survey or a customer survey in which respondents express their intensity of feeling that depend on some factors that can be measured and a few unobservable factors represented by ε. An ordinal scale of say 1-5 represents a spectrum of subjective feeling with 1 implying worst (or strong disagreement) and 5 proxying for best (or strong agreement). The respondents are likely to choose the cell most closely representing their feeling or perception on a certain question. It is assumed that ε is normally distributed with an expected value of zero and variance of unity.

One has the following probabilities:

$$\begin{aligned} &\text{Prob(y=0|x)}=\Phi(-x'\beta) \quad (\mu_0 = 0) \\ &\text{Prob(y=1|x)}=\Phi(\mu_1 - x'\beta) - \Phi(-x'\beta) \\ &\text{Prob(y=2|x)}=\Phi(\mu_2 - x'\beta) - \Phi(\mu_1 - x'\beta) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\text{Prob(y=J|x)}=1-\Phi(\mu_{J-1} - x'\beta) \end{aligned} \tag{3}$$

For all the probabilities to be positive, one must have

$$0 < \mu_1 < \mu_2 < \dots\dots\dots < \mu_{J-1} \qquad (4)$$

As usual the marginal effects of the independent variables *x* on the probabilities are not equal to the coefficients. It is helpful to consider a simple example. Suppose there are five categories. The model thus has only three unknown threshold parameter (the first unknown threshold parameter is normalised to zero). The three probabilities are:

$$\begin{aligned}
&\text{Prob(y=0|x)}=\Phi(-x'\beta)\\
&\text{Prob(y=1|x)}=\Phi(\mu_1 - x'\beta)-\Phi(-x'\beta)\\
&\text{Prob(y=2|x)}=\Phi(\mu_2 - x'\beta)-\Phi(\mu_1 - x'\beta)\\
&\text{Prob(y=3|x)}=\Phi(\mu_3 - x'\beta)-\Phi(\mu_2 - x'\beta)\\
&\text{Prob(y=4|x)}=1-\Phi(\mu_3 - x'\beta)
\end{aligned}$$

(5)

For these probabilities, the corresponding marginal effects of the changes in the independent variables are:

$$\begin{aligned}
&\frac{\partial \text{Prob(y=0|x)}}{\partial \text{x}}=\phi(-x'\beta)\beta\\
&\frac{\partial \text{Prob(y=1|x)}}{\partial \text{x}}=[\phi(-x'\beta)-\phi(\mu_1 - x'\beta)]\beta\\
&\frac{\partial \text{Prob(y=2|x)}}{\partial \text{x}}=[\phi(\mu_1 - x'\beta)-\phi(\mu_2 - x'\beta)]\beta \qquad (6)\\
&\frac{\partial \text{Prob(y=3|x)}}{\partial \text{x}}=[\phi(\mu_2 - x'\beta)-\phi(\mu_3 - x'\beta)]\beta\\
&\frac{\partial \text{Prob(y=4|x)}}{\partial \text{x}}=\phi(\mu_3 - x'\beta)\beta
\end{aligned}$$

The analytical framework presented above is applied to the data set described in Section 3 to identify the main determinants of perceived teaching 'quality'and subject the results to sensitivity analysis. The discussion on these is deferred until Sections 4 and 5.

3 The Data and an Interpretive Overview

The basic data for this study relate to the Student Evaluation of Teaching (SET) surveys across nine courses that included four large second and two large third level undergraduate courses and three large postgraduate courses in economics at a leading Australian university between 2000 and 2006 years involving more than 2400 students. These are 'official' data. Note that these surveys do not include any factors that relate to student or course attributes. The variable codes and definitions and prior expectations about the direction of relationship with the dependent variable are provided in Table 1.

Table 1: Definitions of Variables and Description of SET Data

Variable Code	Description	Expected relation with *TEVAL*
TEVAL	***Dependent variable***: All things considered how would you rate this lecturer's overall effectiveness as a university teacher? (1 – very poor, 5 – outstanding)	-
Independent variables**:** Instructor and course attributes 1- strongly disagree; 5 – strongly agree		-
ORGANISE	The lecturer produced classes that were well organized	Positive
PRESENT	The lecturer presented material in an interesting way	Positive
FEEDBACK	The lecturer gave adequate feedback on my work	Positive
RESPECT	The lecturer treated students with respect	Positive
KNOWWELL	The lecturer seemed to know the subject well	Positive
ENTHUSM	The lecturer communicated his/her enthusiasm for the subject	Positive
THINKMEM	The lecturer emphasized thinking rather than just memorizing	Positive
EXPLAIN	The lecturer gave explanations that were clear	Positive
CONSULT	The lecturer were available for consultation	Positive
LSKILLS	The lecturer helped to improve my learning skills	Positive
CEVAL	Overall I was satisfied with the quality of the course	Positive
OBJECTIV	The course has fulfilled stated objectives	Positive
WORKLOAD	The workload was appropriate for the credit point value of the course	Positive
ASSESS	Assessment requirements were made clear at the beginning of this course	Positive
GRADATTR	I have achieved the graduate attributes which the course aimed to develop (e.g. oral/written communication, team work, critical thinking, problem solving)	Positive
ADMIN	The course was administered well (e.g., sufficient resources were available when needed).	Positive

Note that information on all the variables do not encompass all years. For example, *OBJECTIV, ASSESS, GRADATTR* and *ADMIN* have much fewer observations than most of the other variables because earlier SET surveys did not include these items. The data do not meet the criterion of strict randomness in the sense that courses were not selected at random. This is because many staff members are sensitive to letting others use their *TEVAL* records for research. Nevertheless, the data used in this study relate to a range of courses – including large second and third level undergraduate and postgraduate courses. These were some of the courses with the maximum degree of diversity in student population. Note also that the university collects *TEVAL* data based on random sampling in the sense that only the students present in the class on the day of the evaluation are able to participate in the process. Every student in the population has an equal chance of being present and participating in the SET process.

Table 2 presents the descriptive statistics including median, mean and mode, and inter-quartile spread. Given the ordinal nature of the data, median and mode, not mean, are the appropriate measures of central tendency[6]. It is clear from Table 2, that the distributions of *TEVAL* and other attributes are skewed to the left implying a heavy concentration in the top end of the 5-point scale. In most cases, the highest point on the scale represents the third quartile (Q_3) while the first quartile (Q_1) without exception is located the 3-4 range. It can also be seen that distributions for the lower undergraduate courses relatively less skewed than the upper undergraduate and postgraduate samples. In general, the summary scores indicate higher rates of satisfaction with postgraduate and higher level undergraduate courses. This can also be seen from the degree of concentration in the 4-5 range of the scale and inter-quartile range.

Table 2: Distribution of Overall Perceived Teaching Effectiveness (*TEVAL*) Scores and Related Attributes Based on Student Evaluation of Teaching (SET) Data

Level	Sample Size	Median (Mean, Mode)	Q_1	Q_3	IQR
TEVAL					
All Courses	2413	4 (3.99; 4)	3	5	2
Undergraduate	1545	4 (3.88; 4)	3	5	2
Level 2	986	4 (3.71; 4)	3	4	1
Level 3	566	4 (4.18; 5)	4	5	1
Postgraduate	868	4 (4.19; 5)	4	5	1
ORGANISE					
All Courses	2464	4 (4.05;4)	4	5	1
Undergraduate	1572	4 (4.04; 4)	4	5	1
Level 2	1006	4 (3.89; 4)	3	5	2
Level 3	566	4 (4.30; 4)	4	5	1
Postgraduate	892	4 (4.07; 4)	4	5	1
PRESENT					
All Courses	2465	4 (3.82; 4)	3	5	2
Undergraduate	1571	4 (3.61; 4)	3	4	1
Level 2	1006	4 (3.39:4)	3	4	1
Level 3	565	4 (4; 4)	3.5	5	1.5
Postgraduate	894	4 (4.18; 5)	4	5	1
FEEDBACK					
All Courses	2345	4 (3.74; 4)	3	5	2
Undergraduate	1475	4 (3.61; 3)	3	4	1
Level 2	964	4 (3.56; 3)	3	4	1
Level 3	511	4 (3.71; 4)	3	4	1
Postgraduate	870	4 (3.97; 5)	3	5	2
RESPECT					
All Courses	2461	5 (4.35; 5)	4	5	1
Undergraduate	1568	4 (4.29; 5)	4	5	1
Level 2	1003	4 (4.22; 5)	4	5	1
Level 3	565	5 (4.42; 5)	4	5	1
Postgraduate	893	5 (4.46; 5)	4	5	1
KNOWWELL					
All Courses	2463	5 (4.42; 5)	4	5	1
Undergraduate	1571	5 (4.36; 5)	4	5	1
Level 2	1005	4 (4.25; 5)	4	5	1
Level 3	566	5 (4.56; 5)	4	5	1
Postgraduate	892	5 (4.53; 5)	4	5	1
ENTHUSM					
All Courses	2464	4 (4.22; 5)	4	5	1
Undergraduate	1571	4 (4.10; 4)	4	5	1
Level 2	1005	4 (3.98; 4)	4	5	1
Level 3	566	4 (4.30; 5)	3	5	2
Postgraduate	893	5 (4.43; 5)	4	5	1
THINKMEM					
All Courses	2463	4 (4.03; 4)	3	5	2
Undergraduate	1571	4 (3.91; 4)	3	5	2
Level 2	1005	4 (3.76; 4)	3	4	1
Level 3	566	4 (4.19; 4)	4	5	1
Postgraduate	892	4 (4.22; 4)	4	5	1

Table 2 continued

Level	Sample Size	Median (Mean, Mode)	Q_1	Q_3	IQR
EXPLAIN					
All Courses	2459	4 (3.84; 4)	3	5	2
Undergraduate	1567	4 (3.66; 4)	3	5	2
Level 2	1004	4 (3.48; 4)	3	4	1
Level 3	563	4 (3.97; 4)	3	5	2
Postgraduate	892	4 (4.16; 5)	4	5	1
CONSULT					
All Courses	2374	4 (4.02; 5)	3	5	2
Undergraduate	1503	4 (4; 5)	3	5	2
Level 2	966	4 (3.97; 4)	3	5	2
Level 3	537	4 (4.05; 5)	3	5	2
Postgraduate	871	4(4.06; 4)	3	5	2
LSKILLS					
All Courses	2432	4 (3.68; 4)	3	4	1
Undergraduate	1550	4 (3.51; 4)	3	4	1
Level 2	992	3 (3.37; 3)	3	4	1
Level 3	558	4 (3.76; 4)	3	4	1
Postgraduate	882	4 (3.99; 4)	3	5	2
CEVAL					
All Courses	2235	4 (3.85; 4)	3	5	2
Undergraduate	1347	4 (4; 5)	3	4	1
Level 2	888	4 (3.56; 4)	3	4	1
Level 3	459	4 (3.76; 4)	3	5	2
Postgraduate	888	4 (4.08; 4)	4	5	1
OBJECTIV					
All Courses	936	4 (3.92; 4)	3	5	2
Undergraduate	610	4 (3.76; 5)	3	4	1
Level 2	481	4 (3.59; 4)	3	4	1
Level 3	129	5 (4.36; 5)	4	5	1
Postgraduate	326	4(4.22; 5)	4	5	1
WORKLOAD					
All Courses	2246	4 (3.83; 4)	3	5	2
Undergraduate	1357	4 (3.71; 4)	3	4.5	1.5
Level 2	894	4 (3.59; 4)	3	4	1
Level 3	463	4 (4; 4)	4	5	1
Postgraduate	889	4 (4.03; 4)	4	5	1
ASSESS					
All Courses	1990	4 (3.91; 4)	3	5	2
Undergraduate	1101	4 (3.77; 4)	3	4.5	1.5
Level 2	642	4 (3.65; 4)	3	4	1
Level 3	459	4 (3.97; 4)	3	5	2
Postgraduate	889	4 (4.08; 4)	4	5	1
GRADATTR					
All Courses	677	4 (3.95; 4)	3	5	2
Undergraduate	352	4 (3.86; 4)	3	4	1
Level 2	224	4 (3.73; 4)	3	4	1
Level 3	128	4 (4.08; 3)	4	5	1
Postgraduate	325	4 (4.05; 4)	4	5	
ADMIN					
All Courses	685	4 (4.14; 4)	4	5	1
Undergraduate	358	4 (4.19; 4)	4	5	1
Level 2	229	4 (4.16; 4)	4	5	1
Level 3	129	4 (4.26; 4)	4	5	1
Postgraduate	327	4 (4.07; 4)	4	5	1

Table 3 presents results of *Z* tests for difference of proportions in the 4-5 range of the scale to see if they differ between programs and between levels within the undergraduate program. Except for *ADMIN*, *CONSULT* and *ASSESS* there appears to be a significant difference in the 4-5 range of the agreement rate between Level 3 and Level 2 courses. The proportions for the Level 3 courses are significantly higher than those for the Level 2 courses. Statistically significant differences exist between the relevant proportions for the undergraduate and postgraduate programs except for *ORGANISE* and *KNOWWELL*.

Table 3: Percentage of Agreement Rates in the 4-5 Range of SET Data by Level and Program: Results of Tests of Differences of Proportions*

Variable	Undergraduate		Z statistic	*p*-value**	Program		Z statistic	*p*-value**
	Level 3	Level 2			Post-graduate	Under-graduate		
TEVAL	80.9	63.9	6.99	0.000	80.8	70.0	5.76	0.000
ORGANISE	87.1	74.3	5.99	0.000	78.4	78.9	0.31	0.763
PRESENT	75.0	52.8	8.67	0.000	78.4	60.7	8.99	0.000
FEEDBACK	58.1	50.5	2.78	0.005	69.0	53.2	7.52	0.000
RESPECT	89.4	83.4	3.21	0.001	90.4	85.6	3.43	0.001
KNOWWELL	92.9	84.9	4.68	0.000	90.0	87.8	1.77	0.077
ENTHUSM	87.1	75.5	5.48	0.000	86.7	79.8	4.28	0.000
THINKMEM	80.9	66.2	6.20	0.000	80.0	71.5	4.66	0.000
EXPLAIN	73.2	58.2	5.92	0.000	78.9	63.3	8.05	0.000
CONSULT	69.8	67.6	0.89	0.372	74.6	68.3	3.21	0.001
LSKILLS	62.9	46.8	6.10	0.000	71.4	52.5	9.16	0.000
CEVAL	71.7	58.4	4.77	0.000	77.8	62.8	7.49	0.000
OBJECTIV	89.9	61.7	6.08	0.000	81.3	67.4	4.53	0.000
WORKLOAD	77.3	58.6	6.85	0.000	76.8	65.0	5.97	0.000
ASSESS	70.4	65.0	1.89	0.059	76.9	67.2	4.78	0.000
GRADATTR	78.1	63.8	2.79	0.005	77.2	69.0	2.40	0.016
ADMIN	86.0	84.2	0.45	0.654	77.4	84.9	2.53	0.011

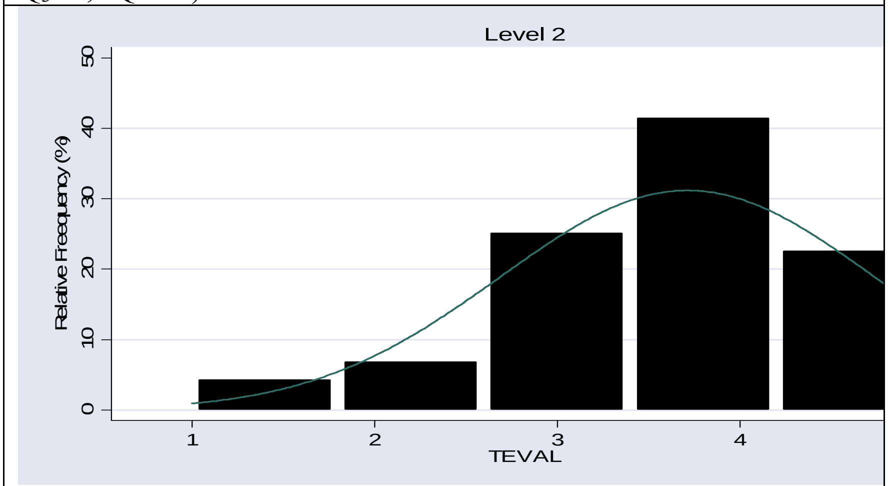
* based on Appendix Table 1; ** for a two-tailed test of significance.

In light of the above, we tested to determine whether the distributions differed between postgraduate and undergraduate samples or between two undergraduate samples. This test was carried out only for *TEVAL*. Visual inspection of the relevant disributions presented in Figure 1 suggests that upper level undergraduate and postgraduate samples have similar distibutions while the lower and upper undergraduate distributions differ. A two-sample Kolmogorov-Smirnov test indicates that *TEVAL* distributions for the two undergraduate samples are significantly different[7]. The same test also revealed that the disributions for Level

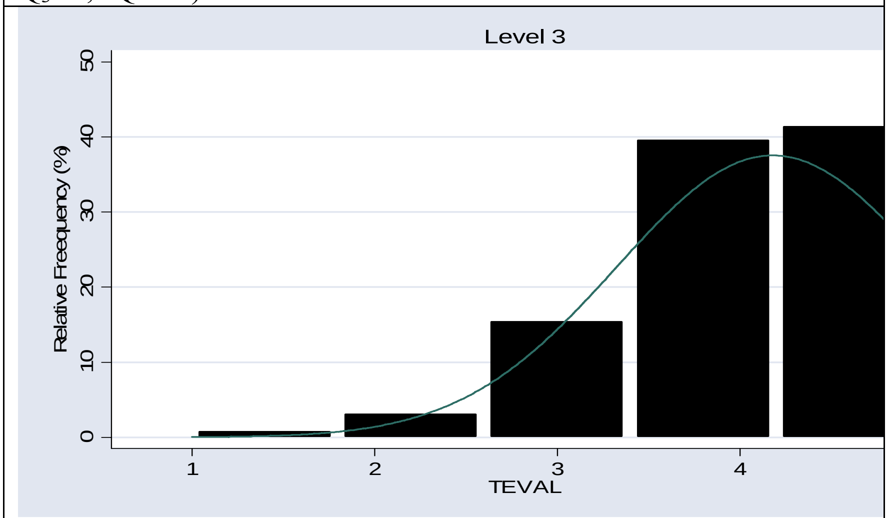
3 undergraduate and the postgraduate samples did not differ significantly[8]. However, as expected the distributions relating the lower undergraduate sample and the postgraduate one were significantly different[9].

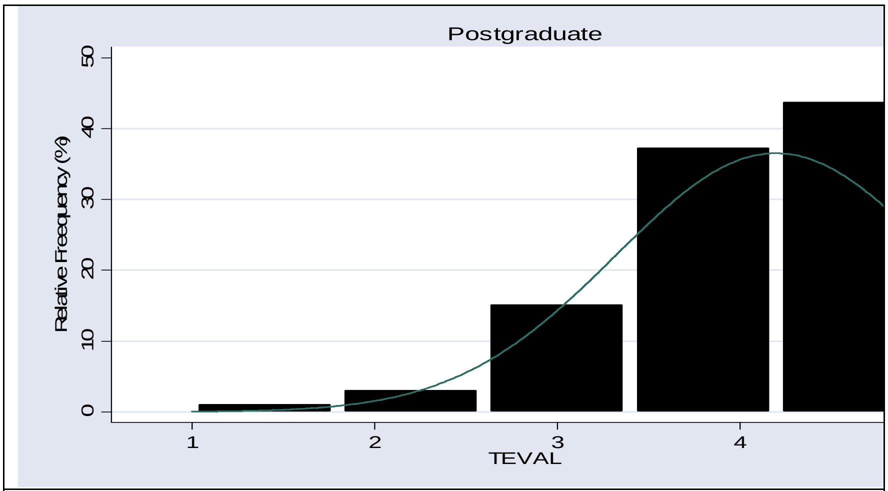
Figure 1: Distribtions of *TEVAL* for two levels of undergrdauate and postgraduate samples.

A: Level 2 Undergraduate Sample (Median=4, Mode=4, Q_1=3, Q_3=4, IQR=1)



B: Level 3 Undergraduate Sample(Median=4, Mode=5, Q_1=4, Q_3=5, IQR=1)





C: Postgraduate Sample (Median=4, Mode=5, Q_1=4, Q_3=5, IQR=1)

Note that as the level of a class increases the apread of the distribution of *TEVAL* scores tends to decline and they tend to move upwards. *Ceteris paribus* this indicates that those who teach lower level classes are likely to obtain lower and more dispersed *TEVAL* scores than those teaching higher level classes. However, at higher levels the differences between those distributions may not be statistically significant.

Several factors may explain this suggested pattern. For example, students in their earlier years may show considerable variation in cottoning on to a new subject. By later years, they are more familiar with its terminology and approach and may show less variation in their degree of learning about the subject. This may be reflected in their *TEVAL* scores. Furthermore, sorting is likley to occur. Those students who are less enthusiastic or less able to cope with a subject are less liklely to continue with it in later years than those who are more capable and enthusiastic. This in all probability willl be reflected in the distribution of the *TEVAL* scores. However, further research is warranted to identify the reasons for the observed changes in the distribution of *TEVAL* with the level of a subject. The results. suggest that the *TEVAL* scores of those teaching to the lower level classes should be adjusted accordingly to be comparable with scores of those teaching higher level courses.

Pearson correlation coefficients between *TEVAL* and the remaining variables are set out in Table 4. The results indicate significant correlations. However, it can also be seen that for the entire sample data *EXPLAIN, PRESENT*, *ORGANISE*, *LSKILLS* and *CEVAL* show the strongest correlation with *TEVAL*. These results are similar to those of Tang (1997). These factors also seem to show similar strengths of correlation with *TEVAL* in the undergraduate and the postgraduate programs.

Table 4: Pearson Correlation Coefficient between *TEVAL* and Other Attributes

Attribute	Correlation with *TEVAL**				
	All courses	Undergraduate			Postgraduate
		Level 2	Level 3	Combined	
ORGANISE	0.704	0.714	0.704	0.724	0.685
N	2411	986	559	1545	866
PRESENT	0.726	0.737	0.721	0.748	0.652
N	2412	986	558	1544	868
FEEDBACK	0.543	0.508	0.553	0.522	0.545
N	2299	948	506	1454	845
RESPECT	0.570	0.571	0.571	0.579	0.530
N	2409	984	558	1542	867
KNOWWELL	0.664	0.677	0.638	0.680	0.611
N	2411	986	559	1545	866
ENTHUSM	0.626	0.596	0.596	0.611	0.624
N	2412	986	559	1545	867
THINKMEM	0.643	0.625	0.632	0.645	0.612
N	2410	985	559	1544	866
EXPLAIN	0.734	0.761	0.735	0.765	0.633
N	2408	986	556	1542	866
CONSULT	0.458	0.407	0.498	0.432	0.514
N	2326	950	531	1481	845
LSKILLS	0.699	0.689	0.666	0.693	0.684
N	2382	974	551	1525	857
CEVAL	0.682	0.647	0.720	0.681	0.651
N	2193	872	459	1331	862
OBJECTIV	0.509	0.432	0.576	0.488	0.514
	895	462	128	590	305
WORKLOAD	0.458	0.352	0.477	0.412	0.501
N	2200	875	462	1337	863
ASSESS	0.490	0.404	0.489	0.444	0.528
N	1947	626	458	1084	863
GRADATTR	0.562	0.555	0.536	0.563	0.553
N	640	209	127	336	304
ADMIN	0.560	0.528	0.658	0.565	0.575
N	647	213	128	341	306

* All the correlation coefficients are significant with a *p*-value = 0.000 for a two-tailed test.

While the above analysis of the salient features of the data and the correlation coefficients are useful we need a more comprehensive analytical framework outlined in Section 2 in order to quantify the effects of changes in the attributes on the *TEVAL* score. This is addressed in Section 4.

4 Empirical Results

Employing the model presented in Section 2 and data described in Section 3, this section presents and analyses empirical results. The dependent variable, *TEVAL*, is coded from 0 to 4. Note that in terms of our model a positive or negative sign of any coefficient implies a higher or lower probability of belonging to the highest category expressing 'strong agreement' or 'best' and a lower or higher probability of belonging the to 'strong disagreement' or 'worst'.

To start with, Table 5 presents the results of probit analysis of SET data using all the attributes. Note that in using all the attributes we have to settle for a set of only 546 observations in total because only these are with all the attributes.

Table 5: Ordered Probit Analysis of Perceived Overall Teaching Effectiveness Score (*TEVAL*) by Perceived Instructor and Course Related Attributes

Variables	All courses	Undergraduate			Postgraduate
		Level 2	Level 3	Combined	
CONSTANT	***5.627 (0.000)	***6.648 (0.000)	***9.439 (0.000)	***7.017 (0.000)	***4.238 (0.000)
ORGANISE	***0.384 (0.000)	0.304 (0.141)	0.277 (0.445)	*0.279 (0.080)	***0.426 (0.003)
PRESENT	***0.404 (0.000)	0.240 (0.139)	***1.070 (0.000)	***0.465 (0.000)	***0.420 (0.004)
FEEDBACK	0.131 (0.114)	0.033 (0.834)	0.128 (0.622)	-0.004 (0.973)	*0.237 (0.066)
RESPECT	0.018 (0.867)	0.099 (0.618)	-0.039 (0.892)	0.092 (0.538)	-0.024 (0.892)
KNOWWELL	0.186 (0.173)	**0.639 (0.013)	0.164 (0.665)	**0.478 (0.012)	-0.129 (0.567)
ENTHUSM	**0.255 (0.034)	0.188 (0.412)	-0.042 (0.904)	0.166 (0.322)	**0.504 (0.011)
THINKMEM	0.055 (0.579)	0.186 (0.347)	-0.300 (0.258)	0.019 (0.894)	0.067 (0.677)
EXPLAIN	***0.236 (0.010)	**0.356 (0.032)	0.383 (0.218)	**0.257 (0.049)	0.174 (0.242)
CONSULT	-0.140 (0.128)	-0.177 (0.333)	0.396 (0.149)	-0.044 (0.745)	-0.199 (0.158)
LSKILLS	***0.310 (0.002)	**0.406 (0.025)	0.453 (0.166)	***0.445 (0.003)	*0.286 (0.059)
CEVAL	***0.812 (0.000)	***0.844 (0.000)	***1.396 (0.000)	***1.026 (0.000)	***0.669 (0.000)
OBJECTIV	-0.118 (0.267)	-0.017 (0.939)	-0.466 (0.106)	-0.167 (0.292)	-0.101 (0.522)
WORKLOAD	0.089 (0.294)	0.122 (0.425)	0.049 (0.859)	0.092 (0.414)	0.043 (0.775)
ASSESS	0.069 (0.411)	-0.204 (0.168)	0.155 (0.475)	-0.036 (0.745)	**0.325 (.030)
GRADATTR	0.059 (0.560)	0.096 (0.620)	-0.149 (0.592)	0.042 (0.777)	0.006 (0.967)
ADMIN	*0.159 (0.087)	-0.076 (0.688)	0.439 (0.164)	0.077 (0.594)	*0.272 (0.061)
μ_1	***1.896 (0.000)	***1.368 (0.000)	***3.615 (0.000)	***1.391 (0.000)	***3.475 (0.000)
μ_2	***4.429 (0.000)	***4.007 (0.000)	***7.448 (0.000)	***4.019 (0.000)	***6.124 (0.000)
μ_3	***6.992 (0.000)	***6.771 (0.000)	ζ	***6.906 (0.000)	***8.576 (0.000)
$\chi^2(10)$	642.97	214.88	148.86	356.01	300.06
N	546	164	113	277	269
Psuedo R^2	0.52	0.52	0.67	0.55	0.52

Notes: *p*-values are in parentheses. ***, **, and * respectively represent 1%, 5% and 10% significant levels for a two-tail test.

ζ For this equation we had only 4 observations with TEVAL = 1 while we use 20 independent variables. This lead to lack of degree of freedoms, and was skipped. Therefore, we have only observations for TEVAL in the 2-4 range. Hence, the number of μ in this case is less than for other equations.

Five estimated regression equations are presented. They relate to: All courses, undergraduate (levels 2 and 3), combined undergraduate and postgraduate. The χ^2 statistics in all models suggest the null hypothesis of all the coefficients equalling zero could be rejected. The values of the pseudo-R^2 range between 0.52 and 0.67 indicating reasonable fits for all the models (Chan et al. 2005, p.30)[10].

The estimated equation shows that for the entire sample only seven attributes (*ORGANISE, PRESENT, ENTHUSM, EXPLAIN, LSKILLS, CEVAL* and *ADMIN*) out of a total of 16 appear to be statistically sifnificant determinants of perceived teaching effectiveness. Of these, the coefficient of *CEVAL* has the highest value (0.812). In all the equations *CEVAL, ORGANISE* and *PRESENT* emerge as significant attributes. In all cases except for the Level 3 undergraduate sample, *LSKILLS* appears significant while *EXPLAIN* is significant in all equations except the ones for the Level 3 undergraduate and the postgraduate courses. *FEEDBACK* appears significant in case of the postgraduate sample while *ENTHUSM* is significant for the entire and the postgraduate samples only. *KNOWWELL* is significant only for the Level 2 and combined undergraduate samples.

In order to maximize the degrees of freedom, the attributes with the lowest number of observations such as *ASSESS, GRADATTR* and *ADMIN* were dropped and all the five equations were re-estimated using the remaining thirteen explanatory variables. These equations are presented in Table 6. For the entire sample, all but *FEEDBACK, ENTHUSIM* and *OBJECTIV* emerged as significant variables. *CEVAL, ORGANISE, PRESENT, EXPLAIN and LSKILLS* were significant in all the estimated equations. However, the orders of magnitudes of their coefficients differed among samples. For instance, in the case of Level 2 undergraduate, *EXPLAIN, CEVAL* and *PRESENT* are the three most important attributes followed by *LSKILLS, KNOWWELL* and *ORGANISE*. For the Level 3 undergraduate sample, the orders of these variables change as do the magnitudes. *CEVAL* and *ORGANISE* are by far the most important variables followed by *EXPLAIN* and *PRESENT*. For the postgraduate sample, the most important attributes are *ORGANISE, CEVAL, LSKILLS*, and *PRESENT*. *FEEDBACK* is not significant in any equation except for the postgraduate sample. *RESPECT* is significant in all except the postgraduate sample. Surprisingly, *THINKMEM* is not a significant factor in the Level 3 undergraduate sample.

Table 6: Ordered Probit Analysis of Overall Perceived Teaching Effectiveness Score (**TEVAL**) by Perceived Instructor and Selected Attributes

Variables	**All courses**	**Undergraduate**			**Postgraduate**
		Level 2	**Level 3**	**Combined**	
CONSTANT	***4.701 (0.000)	***5.547 (0.000)	***5.939 (0.000)	***5.770 (0.000)	***3.543 (0.000)
ORGANISE	***0.438 (0.000)	***0.296 (0.000)	***0.664 (0.000)	***0.381 (0.000)	***0.460 (0.000)
PRESENT	***0.334 (0.000)	***0.460 (0.000)	***0.346 (0.001)	***0.441 (0.000)	***0.219 (0.001)
FEEDBACK	0.065 (0.100)	-0.059 (0.404)	0.150 (0.101)	0.025 (0.646)	***0.171 (0.006)
RESPECT	*0.088 (0.069)	***0.238 (0.003)	*0.200 (0.073)	***0.225 (0.000)	-0.086 (0.272)
KNOWWELL	***0.200 (0.001)	***0.311 (0.001)	0.120 (0.411)	***0.262 (0.001)	0.110 (0.258)
ENTHUSM	0.075 (0.144)	0.062 (0.429)	0.079 (0.499)	0.066 (0.310)	0.130 (0.150)
THINKMEM	***0.145 (0.002)	**0.195 (0.014)	0.078 (0.467)	**0.141 (0.024)	***0.198 (0.010)
EXPLAIN	***0.305 (0.000)	***0.493 (0.000)	***0.414 (0.000)	***0.449 (0.000)	*0.119 (0.073)
CONSULT	**0.091 (.030)	*0.133 (0.058)	***0.246 (0.007)	***0.163 (0.003)	-0.019 (0.787)
LSKILLS	***0.281 (0.000)	***0.317 (0.000)	**0.246 (0.014)	***0.288 (0.000)	***0.285 (0.000)
CEVAL	***0.473 (0.000)	***0.461 (0.000)	***0.689 (0.000)	***0.549 (0.000)	***0.432 (0.000)
OBJECTIV	0.016 (0.70)	-0.002 (0.982)	-0.104 (0.354)	-0.043 (0.453)	0.099 (0.111)
WORKLOAD	*-0.080 (0.063)	**-0.157 (0.018)	0.006 (0.955)	*-0.101 (0.064)	-0.011 (0.888)
μ_1	***1.490 (0.000)	***1.540 (0.000)	***2.260 (0.000)	***1.619 (0.000)	***1.525 (0.000)
μ_2	***3.772 (0.000)	***4.096 (0.000)	***4.972 (0.000)	***4.195 (0.000)	***3.637 (0.000)
μ_3	***5.899 (0.000)	***6.463 (0.000)	***7.586 (0.000)	***6.614 (0.000)	***5.563 (0.000)
$\chi^2(11)$	2328.35	910.02	577.74	1535.69	779.64
N	1830	579	435	1014	816
Psuedo R^2	0.48	0.54	0.55	0.54	0.41

Notes: *p*-values (two-tail) are in parentheses. ***, **, and * respectively represent 1%, 5% and 10% significant levels for a two-tail test.

In light of the above discussion and given that *OBJECTIV* does not feature as a significant variable in any of the estimated equations, it is dropped from subsequent analysis. *WORKLOAD* is also dropped from further analysis as its coefficient presented in Table 5 is difficult to interpret, given the perverse sign. Furthermore, *CEVAL* is dropped as an independent variable[11]. *FEEDBACK* is insignificant for the lower undergraduate class.

Surprisingly, *ENTHUSM* turns out to be an insignificant variable for all the undergraduate samples. Equally surprising is the lack of significance of the coefficient of *CONSULT* for the postgraduate cohort.

Table 7 presents the five estimated equations using all the instructor attributes as independent variables. *RESPECT* is significant only in the undergraduate sample. It is surprising to see that for the postgraduate sample its coefficient has a perverse sign. Lecturer's knowledge about the subject (*KNOWWELL*) does not appear significant for the upper undergraduate and the postgraduate samples.

Table 7: Ordered Probit Analysis of Overall Perceived Teaching Effectiveness Score (*TEVAL*) by Perceived Instructor Attributes and Perceived Course Quality

Variables	All courses	Undergraduate			Postgraduate
		Level 2	Level 3	Combined	
CONSTANT	***4.152 (0.000)	***4.592 (0.000)	***5.297 (0.000)	***4.849 (0.000)	***3.040 (0.000)
ORGANISE	***0.469 (0.000)	***0.392 (0.000)	***0.660 (0.000)	***0.443 (0.000)	***0.458 (0.000)
PRESENT	***0.360 (0.000)	***0.424 (0.000)	***0.411 (0.000)	***0.431 (0.000)	***0.251 (0.000)
FEEDBACK	***0.107 (.002)	0.040 (.469)	***0.235 (.003)	**0.097 (.029)	***0.191 (.002)
RESPECT	0.066 (.121)	**0.126 (.037)	0.147 (.152)	***0.134 (.01)	-0.050 (.512)
KNOWWELL	***0.203 (0.000)	***0.260 (0.000)	*0.240 (.074)	***0.245 (0.000)	0.100 (.292)
ENTHUSM	**0.096 (.032)	0.062 (.313)	0.113 (.286)	0.081 (.129)	**0.187 (.034)
THINKMEM	***0.127 (.002)	**0.120 (.043)	0.092 (.353)	**0.119 (.018)	**0.171 (.022)
EXPLAIN	***0.369 (0.000)	***0.470 (0.000)	***0.457 (0.000)	***0.467 (0.000)	***0.195 (.002)
CONSULT	***0.112 (.002)	***0.145 (.006)	***0.240 (.004)	***0.162 (0.000)	0.013 (.846)
LSKILLS	***0.342 (0.000)	***0.375 (0.000)	***0.268 (.003)	***0.350 (0.000)	***0.378 (0.000)
μ_1	***1.475 (0.000)	***1.451 (0.000)	***2.281 (0.000)	***1.565 (0.000)	***1.450 (0.000)
μ_2	***3.623 (0.000)	***3.770 (0.000)	***4.683 (0.000)	***3.907 (0.000)	***3.380 (0.000)
μ_3	***5.713 (0.000)	***6.034 (0.000)	***7.113 (0.000)	***6.198 (0.000)	***5.222 (0.000)
$\chi^2(10)$	2618.47	1218.52	588.16	1862.20	732.24
N	2242	929	490	1419	823
Psuedo R^2	0.45	0.48	0.51	0.49	0.38

Notes: *p*-values (two-tail) are in parentheses. ***, **, and * respectively represent 1%, 5% and 10% significant levels for a two-tailed test.

The most important factors that impact on *TEVAL* can be identified in order of the magnitudes of their coefficients and are set out in Table 8. These are: *ORGANISE, EXPLAIN, LSKILLS, CONSULT, FEEDBACK,* and *KNOWWELL*. Their impacts, however, vary across samples as can be seen from the information contained in Table 7. In the upper undergraduate, posgraduate as well as the entire samples, *ORGANISE* is the most influential variable affecting *TEVAL* while *EXPLAIN* and *PRESENT* matter more to the lower level undergraduate courses. *CONSULT* does not appear to be important for the postgraduate courses while it is a relatively less important factor for the undergrdaute samples.

Table 8: Six Most Important Factors Influencing *TEVAL* in Order of the Magnitudes of Their Coefficients by Level and Program

Ranking of factors	All courses	Undergraduate			Postgraduate
		Level 2	Level 3	Combined	
1 (Highest)	*ORGANISE* (0.469)	*EXPLAIN* (0.470)	*ORGANISE* (0.660)	*EXPLAIN* (0.467)	*ORGANISE* (0.458)
2	*EXPLAIN* (0.369)	*PRESENT* (0.424)	*EXPLAIN* (0.457)	*ORGANISE* (0.443)	*LSKILLS* (0.378)
3	*PRESENT* (0.360)	*ORGANISE* (0.392)	*PRESENT* (0.411)	*PRESENT* (0.0.431)	*PRESENT* (0.251)
4	*LSKILLS* (0.342)	*LSKILLS* (0.375)	*LSKILLS* (0.268)	*LSKILLS* (0.350)	*EXPLAIN* (0.195)
5	*KNOWWEL L* (0.203)	*KNOWWELL* (0.260)	*CONSULT* (0.240)	*KNOWWELL* (0.245)	*FEEDBACK* (0.191)
6 (Lowest)	*CONSULT* (0.112)	*CONSULT* (0.145)	*FEEDBACK* (0.235)	*CONSULT* (0.162)	*ENTHUSM* (0.187)

This result is consistent with some previous studies such as that of Boex (2000) who found that organisation and clarity are the most important attributes influening the overall teaching effectiveness score. However, Boex relied on used highly aggregated data and it was not clear if the impact of these attributes differed across levels and programs. Our analysis in this paper represents an extension of Boex's study.

5 Results of Sensitivity Analysis

Table 9 and Table 10 present results of sensitivity analysis. They assess the impact on the probability of the *TEVAL* rating of each of the instructor attributes used for the estimated equations in Table 6 increasing from 4 to 5 or decreasing from 4 to 3.

The probability of getting a 5 for teaching effectivensss is most sensitive to *ORGANISE, EXPLAIN, PRESENT*, and *LSKILLS*. For instance, in the case of all courses, increasing the score of *ORGANISE* from 4 to 5, *ceteris paribus* increases the probability of *TEVAL* = 5 from

19.5 per cent to 34.8 per cent. The respective marginal effects of increasing the scores from 4 to 5 *ceteris paribus* in *EXPLAIN, PRESENT* and *LSKILLS* lead to the increases in the probabilities of 31.1, 30.8, and 30.2 per cent in a *TEVAL* score from 4 to 5 from the base level of 19.5 per cent. The remaining variables such as *ENTHUSM*, *THINKMEM*, and *CONSULT* seem to show very low degrees of sensitivity.

Noticeable variation across courses and levels can be identified. A comparison between the estimated equations for the combined undergraduate sample and the one for the postgraduate sample suggests that a transition from 4 to 5 in respect of *ORGANISE* increases the probability of *TEVAL* = 5 from 17.4 per cent to 31.1 per cent in case of the former while chance of getting a 5 increases from 24.7 per cent to 41.1 per cent in case of the latter. Contrasting patterns can also be identified in regard to other important determinants of *TEVAL* such as *PRESENT, FEEDBACK*, *THINKMEM* and *LSKILLS.*

Table 9: Sensitivity Analysis of the Impact on Probability (Measured in Percentage) of *TEVAL* of Rating of Each Selected Attribute increasing from 4 to 5 *ceteris paribus*

All Courses

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	69.6	60.8	63.6	68.4	68.9	66.9
TEVAL=5	19.5	34.8	30.8	22.5	21.3	25.5
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	69.6	68.5	68.1	63.4	68.3	64.0
TEVAL=5	19.5	22.2	23.1	31.1	22.7	30.2

Level 2 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	73.6	67.2	66.5	73.3	72.3	70.1
TEVAL=5	16.6	28.2	29.3	17.6	19.9	23.9
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	73.6	73.0	72.4	65.2	72.0	67.6
TEVAL=5	16.6	18.2	19.8	30.9	20.5	27.6

Table 9 (Continued)
Level 3 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	76.0	60.0	67.7	72.1	73.8	72.0
TEVAL=5	16.9	38.3	29.3	23.5	20.9	23.7
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	76.0	74.4	74.8	66.5	72.0	71.4
TEVAL=5	16.9	20.0	19.4	30.9	23.7	24.5

Combined Level 2 & Level 3 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	73.8	65.3	65.6	72.6	72.1	70.1
TEVAL=5	17.4	31.1	30.7	20.1	21.1	24.5
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	73.8	72.8	72.3	64.6	71.6	67.7
TEVAL=5	17.4	19.6	20.7	31.9	21.9	27.9

Postgraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	63.0	53.6	58.8	60.0	63.5	61.6
TEVAL=5	24.7	41.1	33.3	31.2	23.2	28.0
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*LSKILLS*
TEVAL=4	63.0	60.1	60.4	59.9	62.8	55.8
TEVAL=5	24.7	31.0	30.4	31.3	25.2	38.0

* All attributes =4.

If one compares the two equations relevant to third level undergraduate and postgraduate courses, similar types but differing degrees of sensitivity can be noted in respect of *ORGANISE, PRESENT*, *FEEDBACK* and *LSKILLS*. Contrasting patterns can also be observed between the lower and upper undergraduate courses in respect of sensitivities to changes in these parameters. For example, the sensitivity to the increase in the score for *ORGANISE* from 4 to 5 increases the probability of *TEVAL* rating of 5 more than doubles from 16.9 per cent to 38.3 per cent in case of the third level course while it only increases to 28.2 per cent from 16.6 per cent in case of the second level course.

Likewise one can observe varying types and degrees of sensitivity of probability of *TEVAL* to a transition from a rating of 4 to a rating of 3 in respect of some of the above variables. For instance, for the postgraduate course in order of sensitivity from the highest to the lowest of the four variables are: *ORGANISE, LSKILLS, FEEDBACK* and *EXPLAIN.* On the other hand, for the second level undergraduate course, the four most sensitive variables appear to be: *EXPLAIN, PRESENT, ORGANISE* and *LSKILLS* while those for the third level undergraduate course are: *ORGANISE*, *EXPLAIN*, *PRESENT* and *LSKILLS*.

Table 10: Sensitivity Analysis of the Impact on Probability (measured in percentage) of *TEVAL* of Rating of Each Selected Attribute decreasing from 4 to 3 *ceteris paribus*

All Courses

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	69.6	68.5	69.6	70.2	70.1	70.4
TEVAL=5	19.5	9.2	11.1	16.7	17.7	14.4
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	69.6	70.2	70.7	69.6	70.3	69.8
TEVAL=5	19.5	16.9	16.2	10.9	16.5	11.4

Level 2 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	73.6	73.0	72.6	73.9	74.2	74.0
TEVAL=5	16.6	8.7	8.2	15.6	13.7	10.9
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	73.6	74.0	74.2	72.0	74.2	73.2
TEVAL=5	16.6	15.1	13.8	7.5	13.2	8.9

Level 3 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	76.0	73.9	77.0	77.5	77.3	77.6
TEVAL=5	16.9	5.3	8.6	11.7	13.5	11.6
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	76.0	77.1	76.9	76.7	77.6	77.6
TEVAL=5	16.9	14.2	14.7	7.9	11.6	11.0

Combined Level 2 & Level 3 Undergraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	73.8	73.5	73.6	74.5	74.7	74.8
TEVAL=5	17.4	8.4	8.6	15.1	14.2	11.9
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*SKILLS*
TEVAL=4	73.8	74.4	74.6	73.2	74.7	74.3
TEVAL=5	17.4	15.5	14.6	8.0	13.6	9.9

Table 10 (Continued)
Postgraduate

Probability	Base case*	*ORGANISE*	*PRESENT*	*FEEDBACK*	*RESPECT*	*KNOWWELL*
TEVAL=4	63.0	63.2	64.3	64.3	62.3	63.8
TEVAL=5	24.7	12.7	17.5	19.1	26.3	21.7
Probability	Base case*	*ENTHUSM*	*THINKMEM*	*EXPLAIN*	*CONSULT*	*LSKILLS*
TEVAL=4	63.0	64.2	64.2	64.3	63.1	63.8
TEVAL=5	24.7	19.2	19.7	19.0	24.3	14.4

* All attributes =4.

6 What the Set Data do not Reveal

In light of the discussion in the preceding sections, it can be surmised that the there are a number of critically important issues in measuring teaching effectiveness that the SET data are unable to address.

The purpose of a SET survey is unclear. Does it measure the 'quality'of teaching or does it reflect the impressions or perceptions of students about teaching? If it is the latter, as most likely is the case, then it may be more a subjective measure than an objective measure. As Judge, Hill, Griffiths and Lütkephal and Lee (1988, p.582) put it:

> 'In some cases in empirical analysis the variables we measure are not really what we want to measure. … The proxy variables may be subject to large measurement errors. Even for the observable variables the data may be subject to a variety of errors. Errors may be introduced by the wording of the survey questionnaires. Words such as weak and strong may imply different things to different respondents.'

SET surveys do not and cannot by themselves indicate to lecturers how their *TEVAL* score could be efficiently increased and, therefore, the (supposed) 'quality'of their teaching improved. This leaves the lecturer unable to judge how much effort to put into improving each of the (presumed) explanatory variables assessed in the survey.

Little attention is paid to determining which factors are important explanatory variables. The decision to include variables seems to be more of an administrative one than a scientific choice. Therefore, as found in this article, several variables display a high degree of multicollinearity and others virtually have no influence or no significant influence on *TEVAL* scores. For the entire sample, the coefficients of correlation between an important explanatory variable *EXPLAIN* with other explanatory variables such as *ORGANISE, PRESENT, THINKMEM,* and *LSKILLS* were found to be of the order of 0.620, 0.707, 0.611, and 0.630 respectively for the entire sample. Similar degrees of linear dependency were found to exist between pairs of independent variables at disaggregated levels.

A consequence of multicollinearity is that it may lead to three forms of specification errors in single equation models (Deegan 1976, pp.237-38). These are:

1. Type A - Oversepcification that refers to a situation when a specified model correctly includes all of the variables in the unknown but true model but at the same time incorrectly includes other independent variables not in the true model.

2. Type B - Underspecification that refers to the situation when the postulated model does not include one or more of the regressors that appear in the unknown but true model. This results in biased estimates of parameters.

3. Type C – Under- and overspecification that refers to a situation when the specified model includes independent variables not in the true model but excludes some independent variables found in the true mdel. This type of error leads to biased parameter estimates for entire model.

It is only by carrying out the type of analysis done here that a more scientific choice can be made and the survey results given practical significance[12]. Assuming that the *TEVAL* score identifies all the relevant explanatory variables that impinge on the *TEVAL* score, the analysis conducted here helps to reveal how students construct their evaluation, that is, how much weight they give to each of the explanatory variables.

This raises the question of whether those expected weights are appropriate. Should the wieghts vary across levels or courses at the same levels?. Should weights vary across disciplines? Lecturers/scholars or even the adminsistrators may believe that different weights would be more relevant or appropriate. For example, the latter may believe that *THINKMEM* (emphasis on thinking rather than memorizing, should be given considerable weight. For instance, in some universities one of the central graduate attributes is development of analytical abilities and critical judgment.

However, this study shows that students do not, on average, give it much weight. In fact, the results suggest that on average student looks for certainty and security rather than challenges in their subjects. It is possible that clear pathways, without much critical analysis, and the provision of pointers about how to learn the presented material readily could be preferred by the majority of students. If this is so, SET procedures may favour non-academic types of teaching that involve less critical analysis than many academics consider desirable. A corollary of this could be that this method of teaching evaluation reduces emphasis on reading and consideration of competing intellectual points of view. This could reduce the intrinsic quality of university courses.[13, 14]

The specification of *TEVAL* is vague. The student is asked to rate the "teacher's overall effectiveness as a university teacher". However, nothing is said about effectiveness in what regard. Different criteria may be used by different students. This likely to generate errors in data measurement. As Griliches (1974, pp.973-74) notes that errors in data measurement arise because of the: (1) separation of the data collection and the analysis processes[15]; (2) fuzziness what is it one would like to measure; (3) complexity of the phenomena that one is trying to measure. On all of these grounds the SET prcedure produces data with significant erros in measurement.

Some may judge effectiveness on superficial grounds – whether they were entertained, for example[16]. A more appropriate measure of effectiveness would be a measure of what the students learned or how much their understanding of the focal subject was advanced by the lecturer. This (in some cases) may not correlate highly with presenting a subject in an interesting way. Thus SET procedures could lead to superficial presentation of lectures. Note, however that interesting presentations have a significant impact on *TEVAL* scores. Nevertheless one does not learn what it is that makes the teaching interesting and how well that variable relates to academic achievement which is somewhat wider than learning and difficult to measure accurately.

The scores provided by SET data are averages. The distribution of those scores and what influences those, would be worthy of consideration; for example, a teacher may be highly rated by one group and not by another. How could this be rectified? Is it the quality or the nature of the subject rather than the quality of teaching affecting the score? No constructive use of SET data of this type appears to be made. Judgements based on the results may be superficial, as a result, and even erroneous if the mean *TEVAL* measure of central tendency only is used for assessment. In its present form the SET procedure treats as though the distribution of scores as unimodal which may be unrealistic.[17]

The process of averaging implies that each student in the sample is given equal weight even when some students are much better informed, intellectually superior, and less inclined to be lured by superficial treatment of the subject matter and more interested in the substance than appearance than those coming from the other end of the spectrum. In a recent study Felton Mitchell and Stinson (2004, p.106) found that:

> '..Students who voluntarily evaluate their professors' teaching quality in a public forum are significantly affected by how easy the course and how sexy the instructors offering easy courses tend to be rated more highly. Similarly, instructors perceived as sexy tend to receive higher quality scores. The relation between quality and easiness for sexy professors represents the Halo Effect'.

7 Concluding Comments

Employing SET data and probit analaysis this paper finds that instructor's improvement in organization, presentation and explanation, emphasis on critical and analytical ability, positively impact on the perception of teaching effectiveness. The converse also appears to hold. The impacts of these factors vary between postgraduate and undergraduate pograms as well as between levels within the undergraduate program. Furthermore it was found that scores tend to be systematically influenced by whether the subject is at a lower level or not.

Pragmatic implications of SET procedures are discussed. It is argued that while they are simple to apply, there are dangers of using such an indicator to judge the quality of teaching. In the absence of proper weighting of independent variables or wighting of student quality by some attributes such as their student habits, their effort level and intellectual capabilities, the use of *TEVAL* essentially reduces much of its real significance. The present paper concurs with the view expressed by Sproule (2002, p.288) that in the absence of data on these or suitable proxies, the hypothesis that a particular lecturer has failed or succeded in his/her pedagocial responsibilities remains underdetermined and accepting or rejecting the hypothesis on the basis of a certain mean value of *TEVAL* score is little more than promoting what Radner and Radner (1983) calls pseudoscience (see also Sproule 2002, p.288; Krautmann and Sander 1999;). Thus, as Engdahl, Keating and Perrachoine (1993, p.174) put it:

> '.. The process of student evaluation of their professors is fraught with all of the normal problems of performance evaluation, particularly by nontrained raters, forms/scales construction and validity, construct validity, halo effects, recency effects, central tendency effects…. However, the real problem with the student evaluation is the use of the information for decision making about professors' careers'.

Wright (2006) provides an excellent summary of the concerns raised in the literature and seems to confirm most of the pitfalls of the SET procedures raised in this study.

One of the important criticisms in this paper is the failure to use *TEVAL* data to 'instruct' lecturers on effective means (action) to increase their *TEVAL* scores. A further major problem is that indirectly these scores may undermine the quality of course procedures and may encourage 'spoon-feeding' and reduce 'independent work' by students or fail to maintain let alone enhance academic standards the very thing that a university system should foster.

This paper views SET process to be seriously flawed as an instrument for judging the quality teaching and learning outcomes. It is also open to abuse as noted by Becker (2000, p.114 footnote 4):

> 'End of term student evaluations of teaching may be widely used simply because they are inexpensive to administer, especially when done by student in class, with paid staff involved in processing of the results which is the typical routine followed by departments of economics (Becker and Watts 1999). Less than scrupulous administrators and faculty committees may also use them because they can be dismissed or finessed as needed to achieve desired personnel ends while still mollifying students and giving them a sense of involvement in personnel matters.'

The SET procedure does not tell:

- How much the students have learnt and the quality of that learnt. Many researchers from within and outside the US conducted about 2000 studies which show positive association between different aspects of teaching and student performance in multiple choice tests (see, for example, Wilson 1998). However, the problem with multiple choice tests is that they do not really test the analytical and critical abilities and ability to assess. They can favor rote learning. This apart, as Becker and Watts (1999, p.344) point out that these correlations lay in the 0.2-0.7 range. 'Student evaluation scores explain far less than 50 per cent of the variability in other teaching outcomes, such as test scores, scores from trained classroom observers, alumni surveys, and so on' (Becker 2000, p.2000).
- Whether the lecturer developed the critical and analytical ability of the students regarded as a key learning outcome. As Stapleton and Murkison (2001, pp.289-90) suggest that

  > '.. It is possible for some percentage of faculty members to lower homework requirements and grading standards to increase expected grades production and to increase their instructor excellence scores and learning production scores on some evaluations; and conversely, it is possible for some percentage of faculty members to lower their instructor excellence scores on some student evaluations by increasing homework requirements, raising grade standards, and lowering expected grades'.

- The extent to which the perceived quality of teaching in prerequisite courses have any bearing on the perception of teaching quality in subsequent courses. Perusal of narrative comments often portrays the students' feelings about a sub-discipline. 'I hate (love) microeconomics or macroeconomics etc.' In addition, lecturers in a course often have to build on the prerequisites in which the students might be inadequately prepared for higher-level courses. As a result, the average student may find the subsequent higher-level course(s) too difficult and this may result in the lecturer getting a poor rating.
- On their own, SET scores do not tell how teachers could efficiently increase their *TEVAL* score. One requires sensitivity analysis for this, as is done in this paper. Furthermore, a *TEVAL* score might be more sensitive to the less desirable teaching attributes than more desirable ones. Therefore, the teaching consequences could be unsatisfactory.

As Becker (2000, p.115) notes:

> 'In the 21st century, sole reliance on traditional end-of-term student evaluation of

teaching should not be tolerated. For starters, student evaluations should focus what students know, that is, what they have learnt'.

In our view, Becker assesses these tests from a pragmatic viewpoint. John Dewey, a pragmatic philosopher, argues that truth and valuable scientific methods are things that 'result in successful rules for action' (Stokes, 2003, p131; see also James 1946)[19]. SET data do not result in successful rules for action, in our view. They should be rejected on pragmatic grounds. They should also be rejected on consequential or evolutionary grounds. There is a danger that they will encourage academic institutions to evolve in undesirable academic directions e.g. to prefer rote learning to critical thinking, skepticism and, exposure to a diversity of views. For William James, who influenced John Dewey, 'something is either true or right just in so far as it has successful application to the world' (Stokes, 2003 p.129; see also James 1975). SET data fails the tests of William James.

8. Notes

1. For an excellent summary of the controversy see amongst others, Mason et al. (1995) and Wilson (1998). See also Marsh (1987); Marsh and Roche (1997); Greenwald and Gilmore (1997); d'Appollonia and Abrami (1997); Mckeachie (1997); Becker (2000); Aleamoni (1999); Krautmann and Sander (1999); Gaski (1987) and Wright (2006).
2. Alauddin and Tisdell (2000, p.8) in expressing a similar view stated that ' … the quality of a program lies not necessarily in its immediate high approval rating but in appreciating the quality of value added in terms of analytical abilities of enduring character critical to a variety of situations encountered in a real world context. The real significance of this value added cannot conceivably be appreciated until well after one's completion of the degree and involvement in the workforce. … '. Consequences of asymmetry of information are well discussed in the literature since the pioneering work of Akerlof (1970).
3. Boex (2000) quantified the influence of some of these factors on the overall teaching effectiveness. A large volume of studies including those by Arreola (1995), Centra (1993), Feldman (1976, 1988), and Marsh (1987) defined and measured many of these instructional dimensions. Further details are provided in Boex (2000). See also Alauddin and Butler (2004a, 2004b).
4. Notable exceptions are Mason et al (1995) and subsequently Sproule (2002) who included a range of variables to account for (i) instructor attributes; (ii) student attributes; and (iii) course

attributes. Sproule (2002, p.289) went further in that he mathematically provided the proof for the underdetermination of instructor performance by SET data (see also Laudan and Leplin 1991).

5. This section is adapted from Greene (2000, pp.875-78). Consider a customer survey where responses are coded 1 (worst/strongly disagree), 2, 3, 4 or 5 (best/strongly agree). 'The linear regression model would treat the difference between a 4 and a 3 the same as that between a 3 and a 2, in fact they are only a ranking' (Greene 2000, p.875).

6. The educational literature and the administrators alike routinely use the mean rather than median or mode even though it is patently wrong to do so from a statistical point of view in case of ordinal data. What seems intriguing is that some of the administrators are highly competent mathematicians, statisticians or econometricians who would advise their students to stick to methodological correctness when they teach. However, when wearing the administrators' hat such as head of school, or serving on promotion and tenure committees, they stridently defend the use of mean *TEVAL* score as the indicator of instructors' teaching quality. The heads or other administrators routinely express concern and give warning of failure to uphold (maintain) teaching quality if a staff member records a (mean) score of below 3.5 (on a five-point scale) in any course.

7. The null hypothessis for a no difference between distributions was rejected (Kolmogorov-Smirnov Z statistic = 3.553, p-value =0.000).

8. The null hypothessis for a no difference between distributions could not be rejected (Kolmogorov-Smirnov Z statistic = 0.431, p-value =0.992).

9. The null hypothessis for a no difference between distributions was rejected (Kolmogorov-Smirnov Z statistic = 4.544, p-value =0.000).

10. Since the traditional R^2 is poor measure of goodness of fit because even if a model fits perfectly R^2 will be less than one. Since the model is estimated using a maximum likelihood approach, a pseudo R^2 is defined by McFadden as R^2=1-(L_U/L_R). L_R is the restricted log likelihood, which is the value of the log of the likelihood function at iteration 0 where slope of all parameters are set to zero and L_U is the unrestricted log likelihood, which is the maximized value of log of the likelihood functions. Other choices of pseudo R^2 include the specifications of Cragg-Uhler and Chow (Daykin and Mofafatt, 2002; Greene 2003, p.683).

11. One really wonders whether *CEVAL* should be regarded as an explanatory variable. It could equally be used as a dependent v udged by the content, delivery, both or something else? What weight should be placed on those possibilities?ariable. Furthermore, the term quality of the course is vague. It begs the question of quality in what respect. Is the quality to be judged by the content, delivery, both or something else? What weight should be placed on these possibilities?

12. As the philosopher John Dewey points out an important way to judge methodologies is by their practised consequences.

13. In a recent paper Mason, Steagall and Fabritius (2003) employs a model to predict how grade maximisation as opposed to knowledge maximisation can impact on the course quality. ‘Thus it is clear why students prefer lower quality courses if their goal is to maximise grades, even when holding knowledge constant’, (p.606).

14. ‘To instructors , generating positive student answers to questions about overall effectiveness and comminuation skills may smack of entertainment and dumbing down’ (Becker 2000, p.114). See also McKeachie (1987, p.1219).

15. The collection is largely the responsibility of the organizations such as survey research centers within a university system and may be divorced from the researchers that engage in in-depth analysis of the collected data.

16. The fundamental business dictum that the customer (in our case student) knows best fundamentally alters the teacher-student relationship readily manifesting in ‘consumer satisfaction’ surveys which now form an integral part of the industrial relations domain at the university level (Furedi 2002, pp. 36 ff).The process has led (1) to commodification of education that student increasingly perceive as a commodity for consumption and seek “edutainment” and (2) higher education to enter into new market-oriented forms of relations with their student consumers and the business world (Poynter 2002, p.64). See also Wright (2006, p.418).

17. In one of the courses included in this study, using an identical teaching method two years in row, the same lecturer received significantly different *TEVAL* ratings which dropped from 4.19 in the first year to 3.04 in the second year. The distribution of *TEVAL* score displayed bimodality with just over a third of the sample rating the lecturer in the 1-2 range (very poor to poor) with almost the same proportion rating him in the 4-5 (excellent to outstanding) end of the specturm. A lecturer in a third level economics course got a very poor *TEVAL* score in one year while at the same time being commended for making the most signficant inpact on the students who were doing the same course but were enrolled in degree programs of a non-economics discipline. The distribution of *TEVAL* score for this lecturer displayed bimodality in the preceding years.

18. ‘… Science, both physical and psychological, makes known the condition upon which certain results depends, and therefore puts at the disposal of life a method for controlling them. Psychology will never tell us what to do ethically, nor just how to do it. But it will afford us insight into the conditions which control the formation and and execution of aims, and thus enable human effort to expend itself sanely, rationally and with assurance’ (Dewey, 1963, p.315).

9. Acknowledgements

An earlier version of this paper was presented at the Third Biennial Developments in Business and Economics Education (DEBE) Conference, Cambridge, UK, 1-2 September 2005. We are especially grateful to Chris O'Donnell for his generous assistance with the econometric analysis and to Nghiem Hong Son for his excellent research assistance. We also wish to express our gratitude to our colleagues Rodney Beard, Alan Duhs, Gareth Leeves, Bruce Littleboy and KK Tang for their encouragement and support. Usual *caveats* apply.

10. References

Abrami, P.C. (1989) 'How Should We Use Student Ratings to Evaluate Teaching?', *Research in Higher Education*, **30**(2), pp.221-27.

Akerloff, G. (1970) 'The Market for Lemons: Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics*, **84**, pp.488-500.

Alauddin, M. and Butler, J.E. (2004a) 'From a Vicious Circle of Anxiety to a Virtuous Circle of Learning: Experience of Teaching Statistics to a Heterogeneous Clientele', *American Journal of Applied Sciences*, **1**(3), pp.202-208.

Alauddin, M. and Butler, J.E. (2004b) 'Teaching Economics in a Changing University Environment: Some Australian Experience', *International Journal of Social Economics,* **31** (7-8), pp.706-20.

Alauddin, M. and Tisdell, C.A. (2000) 'Changing Academic Environment and Teaching of Economics at the University-level: Some Critical Issues Analysed with the Help of Microeconomics', *Economic Papers,* **19**(1), pp.1-17.

Aleamoni, L.M. (1999) 'Student Rating Myths Versus Research Facts from 1924 to 1998', *Journal of Personnel Evaluation in Education*, **13**(2), pp.153-66.

Arreola, R.A. (1995) *Developing a Comprehensive Faculty Evaluation System*, Bolton, MA: Anker.

Becker, W.E. (2000) 'Teaching Economics in the 21st Century', *Journal of Economic Perspectives*, **14**(1), pp.109-19.

Becker, W.E. and Watts, M. (1999) 'How Departments of Economics Evaluate Teaching', *American Economic Review*, **89** (May), pp.344-49.

Boex, L.F.J. (2000) 'Attributes of Effective Economics Instructors: An Analysis of Student Evaluations', *Journal of Economic Education*, **31**(4), pp.211-27.

Centra (1993) *Reflective Faculty Evaluation*, San Francisco: Jossey-Bass.

Chan, G., Miller, P.W. and Tcha, M.J. (2005) 'Happiness in University Education', *International Review of Economics Education*, **4**(1), pp.20-45.

Daykin, A. R. and Moffatt, P. G. (2002). ‘Analyzing Ordered Responses: A Review of the Ordered Probit Model’, *Understanding Statistics, 1*(3), 157.

Deegan Jr., J. (1976) ‘The Consequences of Model Missepcification in Regression Analysis’, *MultivariateBehavioral Research*, **11**(2), pp.237-48.

Dewey, J. (1963) *Philosophy, Psychology and Social Practice: Essays Selected, Edited and with a Foreword by Joseph Ratner*, New York: G.P. Putnam’s Sons

Engdahl, R.A., Keating, R.J. and Perrachoine, J. (1993) Effects of Grade Feedback on Srudent Evaluation of Instruction’, *Journal of Management Education*, **17**(2), pp.174-84.

Feldman, K.A. (1976) ‘The Superior College Teacher from the Students’ View’, *Research in Higher Education*, **5**(3), pp.243-48.

Feldman, K.A. (1988) ‘Effective College Teaching from the Students’ and Faculty’s View’, *Research in Higher Education*, **28**(4), pp.291-328.

Felton, J., Mitchell, J. and Stinson, M. (2004) ‘Web-Based Student Evaluations of Professors: The Relations Between Perceived Quality, Easiness and Sexiness’, *Assessment and Evaluation in Higher Education*, **29**(1), pp.91-108.

Furedi, F. (2002) ‘The Bureaucratization of the British University’, in Hayes, D. and Wynyard, R. (eds. 2002) *The McDonaldization of Higher Education*, Wesport, CT. and London: Bergin and Garvey’, pp.33-42.

Greene, W. H. (2003). *Econometric Analysis* (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Griliches, Z. (1974) ‘Errors in Variables and Other Unobservables’, *Econometrica*, **42**(6), pp.971-98.

James, W. (1946) *Pragmatism*: *A New Name for Some Old Ways of Thinking*, New York: Longmans, Green & Co.

James, W. (1975) *The Meaning of Truth*, Cambridge, Mass.: Harvard University Press.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T-C. (1988) *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley.

Krautmann, A. and Sander, W. (1999) ‘Grades and Student Evaluation of Teachers’, *Economics of Education Review*, **18**(1), pp.49-53.

Laudan, L. and Leplin, J. (1991) ‘Empirical Equivalence and Underdetermination’, *Journal of Philosophy*, **88**, pp.449-72.

Marsh, H. (1987) ‘Students’ Evaluation of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research’, *International Journal of Educational Research*, **11**(3), pp.263-388.

Marsh, H. and Roche, L.A. (1997) 'Making Students' Evaluation of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias and Utility', *American Psychologist*, **52**(11), pp.1187-97.

Mason, P.M., Steagall, J.W. and Fabritius, M.M. (1995) 'Student Evaluations of Faculty: A New procedure for Using Aggregate Measures of Performance', *Economics of Education Review*, **14**(4), pp.403-16.

Mason, P.M., Steagall, J.W. and Fabritius, M.M. (2003) 'The Quality of Business Education', *Economics of Education Review*, **22**, pp.603-09.

McKeachie, W. (1997) 'Student Ratings: The Validity of Use', *American Psychologist*, **52**(11), pp.1218-25.

McKlvey, R.D. and Zavoina, W. (1975) 'A Statistical Model for the Analysis of Ordinal Level Dependent Variables', *Journal of Mathematical Sociology*, **4**(1), pp.103-20.

Poynter, G. (2002) 'Modules and Markets: Education and Work in the "Information Age"', in Hayes, D. and Wynyard, R. (eds. 2002) *The McDonaldization of Higher Education*, Wesport, CT. and London: Bergin and Garvey', pp.57-70.

Ritzer, G. (1996) 'McUniversity in the Postmodern Consumer Society', *Quality in Higher Education*, **2**, pp.185-99.

Rodin, M. and Rodin, B. (1972) 'Student Evaluations of Teaching', *Science*, **177**, pp.1164-66.

Sproule, R. (2002) 'The Underdetermination of Instructor Performance by Data from the Student Evaluation of Teaching', *Economics of Education Review*, **21**, pp. 287-294.

Stapleton, R.J. and Murkison, G. (2001) 'Optimizing the Fariness of Student Evaluations: A Study of Correlations Between Instructor Excellence, Study Poridction and, Learning Production, and Expected Grades', *Journal of Management Education*, **25**(3), pp.269-91.

Stokes, P. (2003) *Philosophy: 100 Essential Thinkers.* Arcturus Publishing, London.

Tang, T. L-P. (1997) 'Teaching Evaluation at a Public Institution of Higher Education: Factors Related to Overall Teaching Effectiveness', *Public Personnel Management*, **26**(3), pp.379-89.

Wilson, R. (1998) 'New Research Casts Doubt on Value of Student Evaluations of Professors', *Chronicle of Higher Education*, (January 16), p.A12.

Wright, R.E. (2006) 'Student Evaluations of Faculty: Concerns Raised in the Literature, and Possible Solutions', College Student Journal, **40**(2), pp.417-22.

ISSN 1444-8890

PREVIOUS WORKING PAPERS IN THE SERIES

ECONOMIC THEORY, APPLICATIONS AND ISSUES

1. Externalities, Thresholds and the Marketing of New Aquacultural Products: Theory and Examples by Clem Tisdell, January 2001.
2. Concepts of Competition in Theory and Practice by Serge Svizzero and Clem Tisdell, February 2001.
3. Diversity, Globalisation and Market Stability by Laurence Laselle, Serge Svizzero and Clem Tisdell, February 2001.
4. Globalisation, the Environment and Sustainability: EKC, Neo-Malthusian Concerns and the WTO by Clem Tisdell, March 2001.
5. Globalization, Social Welfare, Labor Markets and Fiscal Competition by Clem Tisdell and Serge Svizzero, May 2001.
6. Competition and Evolution in Economics and Ecology Compared by Clem Tisdell, May 2001.
7. The Political Economy of Globalisation: Processes involving the Role of Markets, Institutions and Governance by Clem Tisdell, May 2001.
8. Niches and Economic Competition: Implications for Economic Efficiency, Growth and Diversity by Clem Tisdell and Irmi Seidl, August 2001.
9. Socioeconomic Determinants of the Intra-Family Status of Wives in Rural India: An Extension of Earlier Analysis by Clem Tisdell, Kartik Roy and Gopal Regmi, August 2001.
10. Reconciling Globalisation and Technological Change: Growing Income Inequalities and Remedial Policies by Serge Svizzero and Clem Tisdell, October 2001.
11. Sustainability: Can it be Achieved? Is Economics the Bottom Line? by Clem Tisdell, October 2001.
12. Tourism as a Contributor to the Economic Diversification and Development of Small States: Its Strengths, Weaknesses and Potential for Brunei by Clem Tisdell, March 2002.
13. Unequal Gains of Nations from Globalisation by Clem Tisdell, Serge Svizzero and Laurence Laselle, May 2002.
14. The WTO and Labour Standards: Globalisation with Reference to India by Clem Tisdell, May 2002.
15. OLS and Tobit Analysis: When is Substitution Defensible Operationally? by Clevo Wilson and Clem Tisdell, May 2002.
16. Market-Oriented Reforms in Bangladesh and their Impact on Poverty by Clem Tisdell and Mohammad Alauddin, May 2002.
17. Economics and Tourism Development: Structural Features of Tourism and Economic Influences on its Vulnerability by Clem Tisdell, June 2002.
18. A Western Perspective of Kautilya's Arthasastra: Does it Provide a Basis for Economic Science? by Clem Tisdell, January 2003.
19. The Efficient Public Provision of Commodities: Transaction Cost, Bounded Rationality and Other Considerations.
20. Globalization, Social Welfare, and Labor Market Inequalities by Clem Tisdell and Serge Svizzero, June 2003.
21. A Western Perspective on Kautilya's 'Arthasastra' Does it Provide a Basis for Economic Science?, by Clem Tisdell, June 2003.
22. Economic Competition and Evolution: Are There Lessons from Ecology? by Clem Tisdell, June 2003.

23. Outbound Business Travel Depends on Business Returns: Australian Evidence by Darrian Collins and Clem Tisdell, August 2003.
24. China's Reformed Science and Technology System: An Overview and Assessment by Zhicun Gao and Clem Tisdell, August 2003.
25. Efficient Public Provision of Commodities: Transaction Costs, Bounded Rationality and Other Considerations by Clem Tisdell, August 2003.
26. Television Production: Its Changing Global Location, the Product Cycle and China by Zhicun Gao and Clem Tisdell, January 2004.
27. Transaction Costs and Bounded Rationality – Implications for Public Administration and Economic Policy by Clem Tisdell, January 2004.
28. Economics of Business Learning: The Need for Broader Perspectives in Managerial Economics by Clem Tisdell, April 2004.
29. Linear Break-Even Analysis: When is it Applicable to a Business? By Clem Tisdell, April 2004.
30. Australia's Economic Policies in an Era of Globalisation by Clem Tisdell, April 2004.
31. Tourism Development as a Dimension of Globalisation: Experiences and Policies of China and Australia by Clem Tisdell, May 2004.
32. Can Globalisation Result in Less Efficient and More Vulnerable Industries? by Clem Tisdell, October 2004.
33. An Overview of Globalisation and Economic Policy Responses by Clem Tisdell, November 2004.
34. Changing Abundance of Elephants and Willingness to Pay for their Conservation by Ranjith Bandara and Clem Tisdell, December 2004.
35. Economic Globalisation: The Process and its Potential Social, Economic, and Environmental Impacts by Clem Tisdell, October 2005.
36. Introduction: An Overview and Assessment of The Economics of Leisure by Clem Tisdell, November 2005.
37. Globalisation and the Economic Future of Small Isolated Nations, Particularly in the Pacific by Clem Tisdell, November 2005.
38. Business Partnerships in a Globalising World: Economic Considerations by Clem Tisdell, December 2005.
39. Economic and Business Relations Between Australia and China: An Overview and an Assessment by Clem Tisdell, November 2006.
40. China's Economic Performance and Transition in Relation to Globalisation: From Isolation to Centre-Stage? by Clem Tisdell, November 2006.
41. Knowledge and the Valuation of Public Goods and Experiential Commodities: Information Provision and Acquisition by Clem Tisdell, November 2006.